September 5, 2025

Randomization and ambiguity perception

YUTARO AKITA
Department of Economics, Pennsylvania State University
ytrakita@gmail.com

KENSEI NAKAMURA
Graduate School of Economics, Hitotsubashi University
kensei.nakamura.econ@gmail.com

We provide a model of preferences over lotteries of acts in which a decision maker behaves as if optimally filtering her ambiguity perception. She has a set of plausible ambiguity perceptions and a cost function over them, and chooses multiple priors to maximize the minimum expected utility minus the cost. We characterize the model by axioms on attitude toward randomization and its timing, uniquely identify the filtering cost from observable data, and conduct several comparatives. Our model can explain Machina's (2009) two paradoxes, which are incompatible with many standard ambiguity models.

KEYWORDS. Ambiguity perception; randomization; information acquisition; cognitive optimization; Machina's paradox.
JEL CLASSIFICATION. D81.

1. INTRODUCTION

1.1. **Outline.** Thought experiments introduced by Ellsberg (1961) suggest that ambiguity influences decision making. The decision maker (DM) usually avoids actions whose outcomes depend on ambiguous events. This avoidance is incompatible with having a single probabilistic belief. Rather, under ambiguity, the DM considers multiple probabilities plausible, which reflects the lack of information. At the same time, the DM may be able to filter her perception of ambiguity to improve the decision. The filtering can be (physically or mentally) costly, so she would avoid unnecessary filtering.

To be concrete, consider an entrepreneur choosing a business plan. She perceives ambiguity regarding several factors that may affect the profitability of each plan (e.g., the future economic trend, actions by rival companies, government policies, etc.). She can invest time and money in research on those factors to make a more precise prediction. The entrepreneur, therefore, must take into account the tradeoff between the potential benefits of a clearer forecast and the costs of obtaining it.

Alternatively, consider an investor choosing between domestic and foreign assets. The foreign asset is unfamiliar to the investor, but has the potential for a higher return. She tries

We are grateful to Yuhta Ishii, Shaowei Ke, Norio Takeoka, and Shohei Yanagita for their helpful comments.

to adopt an advantageous view on the foreign economic situation to mitigate the fear of the unknown. The investor, in this case, adjusts the perception of fearful uncertainty at a mental cost.

This paper proposes a decision-making model under ambiguity, the **costly ambiguity perception** model, that captures the examples above. Ambiguity perceptions are represented by sets of probabilities, and are associated with costs. Given an ambiguity perception, the DM evaluates each prospect according to the maxmin expected utility (MEU) model of Gilboa and Schmeidler (1989); that is, the value of a prospect is the minimum expected utility among the priors in the perception. The DM chooses her ambiguity perception to maximize the associated MEU value minus the cost.

Important behavioral implications of our model are on the attitude toward randomization and its timing. First, the DM favors randomization happening after the state is realized rather than before. To illustrate, consider a box containing 100 balls, each of which is either red or blue with unknown composition. Let f_{red} and f_{blue} be the bets on the draw of the corresponding color, whose stakes are \$100 and \$0. Take two lotteries that differ in the timing of randomization. The first lottery p randomizes the prizes of the bets ex post (after drawing a ball): it pays \$100 or \$0, each with probability half, regardless of the color drawn. Since ex post randomization reduces ambiguity to risk, the DM can hedge against uncertainty. In the second lottery P, the randomization takes place ex ante (before drawing a ball): it is the half-half lottery over f_{red} and f_{blue}. After resolving the randomization in P, the DM always faces an ambiguous bet, so P provides no hedge. By this difference in the hedging power, it is conceivable that the DM prefers p (ex post randomization) to P (ex ante randomization).

Second, ex ante randomization can never be beneficial. Let A be the event that the temperature at Athens is greater than or equal to 20°C, and let B be the corresponding event for the temperature at Beijing. For each $E \in \{A, B\}$, let f_E be the bet on E whose stakes are \$100 and \$0, and let $\$c_E$ be the DM's certainty equivalent of f_E. Also, consider the certainty equivalent $\$c_Q$ of the half-half lottery Q over f_A and f_B. When facing f_A (resp. f_B), the DM's task to filter her ambiguity perception is only on the temperature at Athens (resp. Beijing). Instead, to evaluate Q, she has to engage in filtering on both Athens and Beijing. Thus, ex ante randomization makes the filtering task more difficult, which leads to $c_Q \leq \max\{c_A, c_B\}$. Those two properties are summarized as main axioms in our characterization.

Section 2 introduces the formal model. To model the two types of randomization (ex ante and ex post), we follow the approach of Anscombe and Aumann (1963) and Ke and Zhang (2020). We take as the primitive a preference over lotteries of acts (functions that associate a risky prospect with an uncertain state). The utility representation is introduced in Definition 1. It consists of a von Neumann–Morgenstern (vNM) function, a set of plausible ambiguity perceptions, and a cost function on it, where the last two components constitute the

DM's cost structure of filtering. The utility of each lottery is the maximum of the expected MEU value minus the cost across plausible ambiguity perceptions. We offer several interpretations of the representation, which describe the DM's (physical or mental) process of optimization.

Section 3.1 provides an axiomatic foundation for the costly ambiguity perception model. *Attraction to ex post randomization* and *indifference to mixture timing of constant acts* are axioms that regulate the attitude toward the timing of randomization. They together state that the DM prefers ex post to ex ante randomization when mixing two ambiguous acts, while the timing does not matter when mixing with constant acts. *Ex ante aversion to randomization* and *independence of constant acts* control the attitude toward ex ante randomization. Under them, the DM is averse to ex ante randomization, while she is independent of mixing with constant acts when the mixing weights are kept unchanged. Together with other basic axioms, these axioms characterize the costly ambiguity perception model (Theorem 1).

Section 3.2 examines identification. Our focus is on the uniqueness of the cost structure. Theorem 2 shows that costs are essentially unique in the class of "canonical" cost structures (see Definition 2). Also, we identify the minimum and maximum sets of plausible ambiguity perceptions that represent the same preference. Moreover, we can construct the canonical costs from certainty equivalent data, which are observable in principle.

Section 4 conducts three comparatives, building on the identification result from Section 3.2. Although the choices of ambiguity perceptions are not observable, we can compare them across DMs through the choice data. We show that (i) the more neutral to ex ante randomization a DM is, the more often she changes her choice of ambiguity perceptions; (ii) the more likely she accepts ambiguity, the lower her costs are; (iii) the higher the incentive of filtering she has, the more refined the chosen perceptions are.

Section 5 applies our model to explain the typical choices in Machina's (2009) two examples: the 50–51 and reflection examples. Most of the models of ambiguity-sensitive decision making (the MEU model, the α-MEU model (Ghirardato et al., 2004), the Choquet expected utility model (Schmeidler, 1989), the variational model (Maccheroni et al., 2006), the uncertainty averse model (Cerreia-Vioglio et al., 2011), and the smooth ambiguity model (Klibanoff et al., 2005)) are known to be incompatible with them (Machina, 2009; Baillon et al., 2011). Our numerical examples explain them by allowing the optimal filtering to depend on the act being evaluated.

Finally, Section 6 discusses special cases and related models. Corollary 2 characterizes three special cases, which have been studied in a different framework without ex ante randomization (Payró et al., 2025; Sinander, 2024; Chandrasekher et al., 2022). It clarifies the differences between our model and theirs at a behavioral level. We also consider the "dual" version of our model, in which the DM chooses her ambiguity perception to minimize (instead of maximize) the minimum expected utility at a cost. Corollary 3 characterizes it by replacing *ex ante aversion to randomization* with *ex ante attraction to randomization* in Theorem 1.

All the proofs are relegated to the Appendix.

1.2. **Related Literature.** The idea of costly choice of ambiguity perceptions also appears in an independent work of Payró et al. (2025). They take as the primitive a preference over acts; that is, the domain does not involve ex ante randomization. Since they focus on a different behavioral implication from us—aversion to mixing comonotonic acts (i.e., acts varying in the same direction), each possible ambiguity perception of their DM must be the core of some convex capacity. This restriction does not impact the explainability of Machina's (2009) two examples. Unlike ours, their construction of the representation relies on the properties of comonotonic acts and the Choquet integral. Compared to them, using the richer domain, we provide a sharp identification result and conduct comparatives on the parameters.

Another model in which the DM optimally chooses her ambiguity perception is the dual-self expected utility model by Chandrasekher et al. (2022). They also consider a preference over acts. The key difference from the costly ambiguity perception model is that there is no cost function in their representation. While we interpret the maximization as a single DM's optimal choice of ambiguity perception, their primary interpretation is an intrapersonal game played by two selves: pessimism and optimism. As we show in Appendix C, this model accommodates the reflection example, but is incompatible with the 50-51 example. The latter incompatibility stems from the *certainty independence* axiom, which is central to Chandrasekher et al.'s (2022) characterization.

The "dual" version of the costly ambiguity perception model is a generalization of Ke and Zhang's (2020) double maxmin expected utility model. In their model, the DM chooses her ambiguity perception to minimize the MEU value without costs. Their primitive is the same as ours, but the DM **prefers** ex ante randomization. It captures the DM's belief that even ex ante randomization helps hedge against ambiguity. Despite the similarities in representations, our characterization is not a direct corollary of theirs. We employ a novel technique to derive the cost structure of the representation. For details, see the discussion after Theorem 1.

Taking a preference over menus as the primitive, Ergin and Sarver (2010) and de Oliveira et al. (2017) characterize the behavioral implications of costly information acquisition. In these models, for each decision problem, the DM optimally chooses an information structure to balance the benefits and costs. Our identification result is based on a generalization of de Oliveira et al.'s (2017). Each of the models has been connected to the timing of randomization (Ergin and Sarver, 2015; Pennesi, 2015).

The costly ambiguity perception model is not the only theory that accounts for Machina's two examples. The vector expected utility model (Siniscalchi, 2009) explains them through the DM's aversion to utility variability, which is reflected in the adjustment term. Another approach is to relax the assumption of expected utility on a risky subdomain (Dillenberger and Segal, 2015; Wang, 2022), while our model maintains it. Without assuming any objective randomization, the expected uncertain utility model (Gul and Pesendorfer, 2014) and the two-stage evaluation model (He, 2021) accommodate the examples. The DM with these models evaluates outcomes differently depending on whether the outcomes are on an ambiguous event or not.

2. Model

2.1. Primitives.

Given any set Y, denote by $\Delta_s(Y)$ the set of all finitely supported probability measures on Y; for each $y \in Y$, denote by $\delta[y]$ the Dirac measure at y.

Let Ω be a finite set of **states** with $|\Omega| \geq 2$, and let X be a set of **consequences**. An **act** is a function from Ω to $\Delta_s(X)$. Let $\mathscr{F}$ be the set of all acts. With an abuse of notation, identify each $p \in \Delta_s(X)$ with the constant act of value p. The DM's preference is modeled as a binary relation $\succsim$ on $\Delta_s(\mathscr{F})$. Each member of $\Delta_s(\mathscr{F})$ is called a **lottery**. Denote by $\sim$ and $\succ$ the symmetric and asymmetric parts of $\succsim$, respectively.

For each $(\lambda, (f, g)) \in [0, 1] \times \mathscr{F}^2$, define $\lambda f + (1 - \lambda)g \in \mathscr{F}$ as

$$[\lambda f + (1 - \lambda)g](\omega)(Z) = \lambda f(\omega)(Z) + (1 - \lambda)g(\omega)(Z) \quad \forall(\omega, Z) \in \Omega \times 2^X.$$

Thus, in $\lambda f + (1 - \lambda)g$, a mixture happens **after** a state is realized. Refer to this type of mixture as **ex post randomization**.

For each $(\lambda, (P, Q)) \in [0, 1] \times \Delta_s(\mathscr{F})^2$, define $\lambda P + (1 - \lambda)Q \in \Delta_s(\mathscr{F})$ as

$$[\lambda P + (1 - \lambda)Q](F) = \lambda P(F) + (1 - \lambda)Q(F) \quad \forall F \in 2^{\mathscr{F}}.$$

Thus, in $\lambda P + (1 - \lambda)Q$, a mixture happens **before** a state is realized. Refer to this type of mixture as **ex ante randomization**. In particular, $\lambda\delta[f] + (1 - \lambda)\delta[g]$ is a lottery that yields f and g with probability λ and $1 - \lambda$, respectively. It is distinguished from the degenerate lottery $\delta[\lambda f + (1 - \lambda)g]$ at the ex post randomization of f and g.

2.2. Representation.

Let $\Delta(\Omega)$ be the set of all probability measures on Ω, endowed with the total variation distance. Let $\mathbb{K}$ be the set of all nonempty compact convex subsets of $\Delta(\Omega)$, endowed with the Hausdorff metric. Each member of $\mathbb{K}$ represents an ambiguity perception: the set of probability distributions the DM believes plausible. A **vNM function** is a nonconstant mixture linear real-valued function on $\Delta_s(X)$. With a vNM function u and an ambiguity perception M, the MEU model evaluates each act f as $\min_{\mu \in M} \int u \circ f \, d\mu$.

Our representation captures the DM who optimally filters her ambiguity perception at a cost. The costs are measured by a cost function over feasible ambiguity perceptions. An (extended) real-valued function is **grounded** if its infimum is zero. A **cost structure** is a pair of a nonempty compact subset of $\mathbb{K}$ and a lower semicontinuous grounded function on it.

Definition 1. A **costly ambiguity perception representation of** $\succsim$ is a pair $\langle u, (\mathbb{M}, c)\rangle$ of a surjective vNM function and a cost structure such that the real-valued function U on $\Delta_s(\mathscr{F})$ of the form

$$U(P) = \max_{M \in \mathbb{M}} \left[\int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f) - c(M) \right]$$

represents $\succsim$.

A **costly ambiguity perception preference** is a binary relation on $\Delta_{\mathrm{s}}(\mathscr{F})$ that has a costly ambiguity perception representation.

In a costly ambiguity perception representation $\langle u, (\mathbb{M}, c) \rangle$, the set $\mathbb{M}$ represents the DM's set of plausible ambiguity perceptions, and c does the cost to filter her ambiguity perception to each $M \in \mathbb{M}$. The DM chooses her ambiguity perception at each lottery to maximize the expectation of the MEU value minus the cost. We offer several interpretations of the representation. Under each of the interpretations, the representation can be seen as a reduced form of the DM's process of subjective optimization, which is not directly observable:

(i) Physical information acquisition. The DM acquires outside information to reduce ambiguity. She selects the amount or source of information to balance the benefit and cost.

(ii) Contemplation. The DM contemplates what the "true" probability law is. The contemplation process excludes probability distributions that she considers unreasonable.

(iii) Correction of a bias. The DM is biased to behave robustly, which leads to excessive ambiguity aversion. Aware of such a bias, she tries to correct it by choosing a smaller perception.

(iv) Mitigation of fear. The DM fears the ambiguous environment she faces. She tries to mitigate the fear by choosing a favorable ambiguity perception.

3. Foundations

3.1. **Characterization.** We introduce several axioms on $\succsim$.

We begin with a standard requirement. The relation $\succsim$ is **nondegenerate** if there exists $(P, Q) \in \Delta_{\mathrm{s}}(\mathscr{F})^2$ such that $P \succ Q$; and **mixture continuous** if for each $(P, Q, R) \in \Delta_{\mathrm{s}}(\mathscr{F})^3$, the sets $\{ \lambda \in [0,1] \mid \lambda Q + (1-\lambda) R \succsim P \}$ and $\{ \lambda \in [0,1] \mid P \succsim \lambda Q + (1-\lambda) R \}$ are closed.

Axiom 1 (Regularity)**.** The relation $\succsim$ is nondegenerate, complete, transitive, and mixture continuous.

The following axiom requires monotonicity with respect to the first-order stochastic dominance induced by statewise dominance. Given any binary relation $\succcurlyeq$ on a set Y, a subset Z of Y is **$\succcurlyeq$-increasing** if Z includes $\{ y \in Y \mid y \succcurlyeq z \}$ for each $z \in Z$. Define the relation $\succsim_{\mathrm{d}}$ on $\mathscr{F}$ by $f \succsim_{\mathrm{d}} g$ if $\delta[f(\omega)] \succsim \delta[g(\omega)]$ for each $\omega \in \Omega$. Define the relation $\succsim_{\mathrm{FSD}}$ on $\Delta_{\mathrm{s}}(\mathscr{F})$ by $P \succsim_{\mathrm{FSD}} Q$ if $P(F) \geq Q(F)$ for each $\succsim_{\mathrm{d}}$-increasing subset F of $\mathscr{F}$.

Axiom 2 (FSD)**.** For each $(P, Q) \in \Delta_{\mathrm{s}}(\mathscr{F})^2$, if $P \succsim_{\mathrm{FSD}} Q$, then $P \succsim Q$.

We introduce two axioms on the attitude toward the timing of randomization. First, as discussed in Section 1, while ex post randomization can help the DM hedge against uncertainty, ex ante randomization does not. This difference in hedging power motivates the DM to prefer the former to the latter.

Axiom 3 (Attraction to ex post randomization)**.** For each $((\kappa, \lambda), R, (f, g)) \in [0,1]^2 \times \Delta_s(\mathscr{F}) \times \mathscr{F}^2$, we have $\kappa\delta[\lambda f + (1-\lambda)g] + (1-\kappa)R \succsim \kappa[\lambda\delta[f] + (1-\lambda)\delta[g]] + (1-\kappa)R$.

Second, if an act is mixed with a constant act, even the ex post randomization of them does not provide hedging. Thus, the DM has no reason to prefer ex post randomization, since *attraction to ex post randomization* is motivated by its hedging power. The following axiom states that the mixture timing of constant acts does not matter for the DM.

Axiom 4 (Indifference to mixture timing of constant acts)**.** For each $((\kappa, \lambda), R, f, p) \in [0,1]^2 \times \Delta_s(\mathscr{F}) \times \mathscr{F} \times \Delta_s(X)$, we have $\kappa\delta[\lambda f + (1-\lambda)p] + (1-\kappa)R \sim \kappa[\lambda\delta[f] + (1-\lambda)\delta[p]] + (1-\kappa)R$.

Combined with *attraction to ex post randomization*, this axiom implies that the DM strictly prefers ex post randomization to ex ante randomization only when two acts to be mixed both have ambiguity.

We then proceed to two axioms on the attitude toward ex ante randomization. Not only useless to hedge against uncertainty, ex ante randomization $\lambda P + (1-\lambda)Q$ makes it unable for the DM to tailor her filtering of the ambiguity perception to the lottery P or Q, as illustrated in Section 1. Consequently, ex ante randomization is unfavorable to the DM with optimal filtering.

Axiom 5 (Ex ante aversion to randomization)**.** For each $(\lambda, (P, Q)) \in [0,1] \times \Delta_s(\mathscr{F})^2$, if $P \succsim Q$, then $P \succsim \lambda P + (1-\lambda)Q$.

The idea behind *ex ante aversion to randomization* is close to what motivates the *aversion to contingent planning* axiom (Ergin and Sarver, 2010; de Oliveira et al., 2017) in the context of costly information acquisition: mixture makes the DM's information acquisition problem more difficult. Our argument suggests that a similar intuition goes through even when we adopt interpretations other than costly information acquisition.

The second axiom in this category postulates that the ranking over two lotteries is independent of their common unambiguous parts.

Axiom 6 (Independence of constant acts)**.** For each $(\lambda, (P, Q), (p, q)) \in [0,1] \times \Delta_s(\mathscr{F})^2 \times \Delta_s(X)^2$, if $\lambda P + (1-\lambda)\delta[p] \succsim \lambda Q + (1-\lambda)\delta[p]$, then $\lambda P + (1-\lambda)\delta[q] \succsim \lambda Q + (1-\lambda)\delta[q]$.

For each constant act p, the optimal ambiguity perception at a lottery of the form $\lambda P + (1-\lambda)\delta[p]$ depends only on λ and P, since the perception choice does not affect the value of the constant act p. Thus, we can interpret *independence of constant acts* as requiring that changes in parts irrelevant to filtering do not affect the DM's ranking.

The final axiom ensures that the range of the vNM function is the entire real line.

Axiom 7 (Unboundedness). For each $(p,q) \in \Delta_s(X)^2$ with $\delta[p] \succ \delta[q]$, there exists $(r,s) \in \Delta_s(X)^2$ such that $\delta[\frac{1}{2}r + \frac{1}{2}q] \succsim \delta[p]$ and $\delta[q] \succsim \delta[\frac{1}{2}p + \frac{1}{2}s]$.

The following is the main result of this paper. The axioms introduced in this section are all the behavioral implications of costly ambiguity perception preferences.

Theorem 1. *The relation $\succsim$ satisfies regularity, FSD, attraction to ex post randomization, indifference to mixture timing of constant acts, ex ante aversion to randomization, independence of constant acts, and unboundedness if and only if it has a costly ambiguity perception representation.*

The sufficiency part of the proof of Theorem 1 exploits results from convex analysis and the notion of niveloids developed in Cerreia-Vioglio et al. (2014) (see Appendices A.1 and A.2). To do so, we transform each lottery to a continuous real-valued function on a topological space. We begin by finding a vNM function u that represents the restriction of $\succsim$ to $\Delta_s(X)$. For each $P \in \Delta_s(\mathscr{F})$, denote by $P_u \in \Delta_s(\mathbb{R}^\Omega)$ the pushforward of P under $f \mapsto u \circ f$. Then, we define a topological space $\mathbb{U}$ and a mixture linear mapping $m \mapsto m^\vee$ from $\Delta_s(\mathbb{R}^\Omega)$ to the set of all bounded continuous real-valued functions on $\mathbb{U}$ (see Appendix A.3 for the precise definitions). Our construction suggests that $P_u^\vee = Q_u^\vee$ implies $P \sim Q$. Using *ex ante aversion to randomization* and *independence of constant acts*, we can construct a convex niveloid W on $\{ m^\vee \mid m \in \Delta_s(\mathbb{R}^\Omega) \}$ such that $P \mapsto W(P_u^\vee)$ represents $\succsim$. Thus, by the mixture linearity of $m \mapsto m^\vee$, there exists a pair $(\mathscr{V}, \gamma)$ of a set of real-valued functions on $\mathbb{R}^\Omega$ and a lower semicontinuous real-valued function on it such that W has the form

$$W(m^\vee) = \max_{V \in \mathscr{V}} \left(\int V \, \mathrm{d}m - \gamma(V) \right).$$

An important step is to restrict $\mathscr{V}$ to the class of MEU functions. This step relies on Proposition 6 in Appendix A.2, which is our main technical innovation. It states that we can take $\mathscr{V}$ as the set satisfying

$$\int V \, \mathrm{d}l \geq \int V \, \mathrm{d}\tilde{l} \quad \forall V \in \mathscr{V}$$
$$\iff W(\lambda l^\vee + (1-\lambda)m^\vee) \geq W(\lambda \tilde{l}^\vee + (1-\lambda)m^\vee) \quad \forall (\lambda, m) \in (0,1] \times \Delta_s(\mathbb{R}^\Omega).$$

Combined with *attraction to ex post randomization* and *indifference to mixture timing*, we can show that each member of $\mathscr{V}$ can be written as an MEU function with respect to a closed convex subset of $\Delta(\Omega)$. A related result to Proposition 6 has been shown by Cerreia-Vioglio et al. (2015), but it is not applicable in our setting. Specifically, Proposition 6 extends their result to a convex niveloid on a more general domain—a set with possibly empty interior.

Although Theorem 1 extends Ke and Zhang's (2020) representation theorem of a preference on a similar domain, our proof strategy is different from theirs. Their proof relies on the relationship between $\succsim$ and its subrelation $\succsim^*$, which we define in Section 3.2. They show that $\succsim^*$ has a "multi-MEU representation" and the original relation $\succsim$ can be recovered through a "cautious completion" of $\succsim^*$. This leads to a minimization of MEU functions over candidate ambiguity perceptions without costs. In our case, the argument of cautious completion is not applicable, and we use properties of niveloids to derive a cost structure.

3.2. **Identification.** The restrictions imposed on cost structures in Definition 1 are not enough to identify them. For instance, if a cost structure $(\mathbb{M}, c)$ satisfies $M \subseteq M'$ and $c(M) < c(M')$ for some $(M, M') \in \mathbb{M}^2$, then slightly decreasing the cost $c(M')$ does not change the DM's preference (because she never chooses M' anyway).

The following definition puts further restrictions on cost structures.

Definition 2. A cost structure $(\mathbb{M}, c)$ is **canonical** if

(i) for each $(M, M') \in \mathbb{M}^2$, if $M \subseteq M'$, then $c(M) \geq c(M')$;

(ii) $\mathbb{M}$ and c are convex.

A costly ambiguity perception representation $\langle u, (\mathbb{M}, c)\rangle$ is **canonical** if $(\mathbb{M}, c)$ is canonical.

In Definition 2, condition (i) requires the monotonicity of costs with respect to set inclusion. It means that the additional filtering starting from any ambiguity perception is costly. Without condition (ii), there exists $(\lambda, (M, M')) \in (0,1) \times \mathbb{M}^2$ such that $c(\lambda M + (1-\lambda)M') > \lambda c(M) + (1-\lambda)c(M')$. In this case, the DM never chooses $\lambda M + (1-\lambda)M'$, so slightly decreasing $c(\lambda M + (1-\lambda)M')$ does not affect her preference.

While canonicality helps identify costs, it still does not pin down the DM's set of plausible perceptions. For instance, if the DM chooses some ambiguity perception (say $\Delta(\Omega)$) only when evaluating constant acts, then we can freely exclude it from the set of perceptions. To explore the smallest possible canonical cost structure, we introduce the following definitions. For each binary relation $\succcurlyeq$ on $\Delta_s(\mathcal{F})$, a **multi-MEU representation of** $\succcurlyeq$ is a subset $\mathbb{M}$ of $\mathbb{K}$ such that $P \succcurlyeq Q$ if and only if

$$\int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu\right) \mathrm{d}P(f) \geq \int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu\right) \mathrm{d}Q(f) \quad \forall M \in \mathbb{M}.$$

Define the relation $\succsim^*$ on $\Delta_s(\mathcal{F})$ by $P \succsim^* Q$ if $\lambda P + (1-\lambda)R \succsim \lambda Q + (1-\lambda)R$ for each $(\lambda, R) \in (0,1] \times \Delta_s(\mathcal{F})$.

Proposition 1. *If $\succsim$ is a costly ambiguity perception preference, then $\succsim^*$ has a unique compact convex multi-MEU representation.*

We are now ready to state our identification result. For each vNM function u representing the restriction of a costly ambiguity perception preference $\succsim$ to $\Delta_s(X)$, define $c^{\star}_{\succsim,u} \colon \mathbb{K} \to [0, \infty]$

by

$$c^{\star}_{\succsim,u}(M) = \sup_{P\in\Delta_s(\mathscr{F})}\left[\int\left(\min_{\mu\in M}\int u\circ f\,\mathrm{d}\mu\right)\mathrm{d}P(f) - u(\bar{P})\right],$$

where for each $P \in \Delta_s(\mathscr{F})$, let $\bar{P} \in \Delta_s(X)$ be such that $\delta[\bar{P}] \sim P$. Given any extended real-valued function γ, let dom γ be the inverse image of $[-\infty, \infty)$ under γ. The next theorem shows four important properties of canonical cost structures: (i) every costly ambiguity perception preference has a canonical representation; (ii) the smallest and largest sets of perceptions exist; (iii) the cost function is unique given a vNM function and a set of perceptions; (iv) the unique cost function can be recovered from observable data.

Theorem 2. *Let u be a vNM function representing the restriction of a costly ambiguity perception preference $\succsim$ to $\Delta_s(X)$, and let $\mathbb{M}^*$ be the compact convex multi-MEU representation of $\succsim^*$.*

(i) *$\langle u, (\mathbb{M}, c)\rangle$ is a canonical costly ambiguity perception representation of $\succsim$ if and only if $\mathbb{M}$ is a compact convex subset of $\mathbb{K}$ such that $\mathbb{M}^* \subseteq \mathbb{M} \subseteq \operatorname{dom} c^{\star}_{\succsim,u}$ and $c = c^{\star}_{\succsim,u}|_{\mathbb{M}}$.*

(ii) *$\langle u, (\mathbb{M}, c)\rangle$ is a canonical costly ambiguity perception representation of $\succsim$ and $\mathbb{M}$ is $\supseteq$-increasing if and only if $\mathbb{M} = \operatorname{dom} c^{\star}_{\succsim,u}$ and $c = c^{\star}_{\succsim,u}|_{\mathbb{M}}$.*

The proof of Theorem 2 is based on an intermediate result, Proposition 7, which generalizes Theorem 2 of de Oliveira et al. (2017) to a more abstract setting. The uniqueness of the set of perceptions does not have a counterpart in de Oliveira et al. (2017), since they deal with extended real-valued costs and take as the domain the largest possible set while ours are real-valued.

Immediately from Theorem 2 and the uniqueness of vNM functions, we obtain the uniqueness of costly ambiguity perception representations. We omit its straightforward proof.

Corollary 1. *If $\langle u, (\mathbb{M}, c)\rangle$ and $\langle u', (\mathbb{M}', c')\rangle$ are canonical costly ambiguity perception representations of the same relation and if $\mathbb{M}$ and $\mathbb{M}'$ are $\supseteq$-increasing, then there exists $(\alpha, \beta) \in \mathbb{R}_{++} \times \mathbb{R}$ such that $u' = \alpha u + \beta$, $\mathbb{M} = \mathbb{M}'$, and $c' = \alpha c$.*

4. Comparatives

Consider two DMs 1 and 2. Each DM i has a preference $\succsim_i$ on $\Delta_s(\mathscr{F})$ with a canonical costly ambiguity perception representation $\langle u_i, (\mathbb{M}_i, c_i)\rangle$.

4.1. Tolerance of ex ante randomization.

For each $(i, P) \in \{1, 2\} \times \Delta_s(\mathscr{F})$, let $\bar{P}_i \in \Delta_s(X)$ be such that $\delta[\bar{P}_i] \sim_i P$.

Definition 3. DM 1 is **more tolerant of ex ante randomization than DM 2** if for each $(\lambda, (P, Q)) \in [0, 1] \times \Delta_s(\mathscr{F})^2$,

$$\lambda\delta[\bar{P}_2] + (1-\lambda)\delta[\bar{Q}_2] \sim_2 \lambda P + (1-\lambda)Q \implies \lambda\delta[\bar{P}_1] + (1-\lambda)\delta[\bar{Q}_1] \sim_1 \lambda P + (1-\lambda)Q.$$

Under *regularity* and *independence of constant acts*, the following condition is equivalent to *ex ante aversion to randomization*: for each $(\lambda, (P, Q), (p, q)) \in [0, 1] \times \Delta_s(\mathscr{F})^2 \times \Delta_s(X)^2$, if $P \sim \delta[p]$ and $Q \sim \delta[q]$, then $\lambda\delta[p] + (1 - \lambda)\delta[q] \succsim \lambda P + (1 - \lambda)Q$. Thus, the indifference $\lambda\delta[\bar{P}_i] + (1-\lambda)\delta[\bar{Q}_i] \sim_i \lambda P + (1-\lambda)Q$ suggests that DM i is indifferent to ex ante randomizing P and Q. A DM will exhibit this indifference when her sets of optimal ambiguity perceptions at P and Q share the same perception, which remains optimal after the randomization. Consequently, the attitude toward ex ante randomization should be tied to changes in optimal ambiguity perceptions.

To formalize the intuition, for each $i \in \{1, 2\}$, define the correspondence $\mathscr{C}_i \colon \Delta_s(\mathscr{F}) \rightrightarrows \mathbb{M}_i$ by

$$\mathscr{C}_i(P) = \operatorname*{arg\,max}_{M \in \mathbb{M}_i} \left[\int \left(\min_{\mu \in M} \int u_i \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f) - c_i(M) \right].$$

That is, $\mathscr{C}_i(P)$ is the set of DM i's optimal ambiguity perceptions at P.

Proposition 2. *DM 1 is more tolerant of ex ante randomization than DM 2 if and only if for each $(P, Q) \in \Delta_s(\mathscr{F})^2$,*

$$\mathscr{C}_2(P) \cap \mathscr{C}_2(Q) \neq \emptyset \implies \mathscr{C}_1(P) \cap \mathscr{C}_1(Q) \neq \emptyset.$$

An empty intersection of $\mathscr{C}_i(P)$ and $\mathscr{C}_i(Q)$ indicates that DM i changes her ambiguity perception if the alternative changes from P to Q. Thus, according to the contrapositive of Proposition 2, the more ex ante averse to randomization a DM is, the more often she changes her choice of ambiguity perceptions, and vice versa.

4.2. **Tolerance of ambiguity.** The following definition extends the notion of comparative ambiguity aversion by Ghirardato and Marinacci (2002) to include ex ante randomization.

Definition 4. DM 1 is **more tolerant of ambiguity than DM 2** if for each $(P, p) \in \Delta_s(\mathscr{F}) \times \Delta_s(X)$,

$$P \succsim_2 \delta[p] \implies P \succsim_1 \delta[p].$$

This condition implies that DM 1 is more tolerant of ex ante randomization than DM 2.

A DM accepts an ambiguous lottery over a constant act more often if it becomes less costly to filter her ambiguity perception. Write $u_1 \approx u_2$ if u_2 is a positive affine transformation of u_1.

Proposition 3. *DM 1 is more tolerant of ambiguity than DM 2 if and only if $u_1 \approx u_2$ and $c^\star_{\succsim_1, u_1} \leq c^\star_{\succsim_2, u_1}$.*

By Theorem 2 (ii), for $\supseteq$-increasing $\mathbb{M}_i$'s, Proposition 3 suggests that if DM 1 is more tolerant of ambiguity than DM 2, then $\mathbb{M}_1 \supseteq \mathbb{M}_2$.

4.3. **Filtering incentives.** The following notion strengthens comparative tolerance of ambiguity.

Definition 5. DM 1 **has higher filtering incentives than DM** 2 if for each $(\lambda, (P, Q), p) \in [0, 1] \times \Delta_s(\mathscr{F})^2 \times \Delta_s(X)$,

$$\lambda P + (1-\lambda)Q \succsim_2 \lambda\delta[p] + (1-\lambda)Q \implies \lambda P + (1-\lambda)Q \succsim_1 \lambda\delta[p] + (1-\lambda)Q.$$

To evaluate constant acts, ambiguity perceptions are irrelevant. Thus, in the lottery $\lambda\delta[p] + (1-\lambda)Q$, replacing $\delta[p]$ with any other lottery P requires additional filtering of ambiguity perceptions, which can be more costly. A DM accepts the replacement if the additional benefits outweigh the costs. Hence, higher filtering incentives should reflect higher benefits from filtering. Define the relation $\trianglerighteq$ on $2^{\mathbb{K}}$ by $\mathbb{M} \trianglerighteq \mathbb{M}'$ if for each $M' \in \mathbb{M}'$, there exists $M \in \mathbb{M}$ such that $M \subseteq M'$.

Proposition 4. *DM 1 has higher filtering incentives than DM 2 if and only if $u_1 \approx u_2$ and $\mathscr{C}_1(P) \trianglerighteq \mathscr{C}_2(P)$ for each $P \in \Delta_s(\mathscr{F})$.*

5. Application

5.1. The 50–51 example. Consider Machina's (2009) thought experiment with a box containing 101 balls, called the 50–51 example. Out of the balls, 50 are either red or blue, and 51 are either green or purple. A ball is drawn at random from the box. The DM is offered four acts f_1, …, f_4. Each act pays off according to the color of the drawn ball, as described in Table 1. For simplicity, the payoffs are measured in units of utility.

	50 balls		51 balls	
	Red	Blue	Green	Purple
f_1	200	200	100	100
f_2	200	100	200	100
f_3	300	200	100	0
f_4	300	100	200	0

TABLE 1. The 50–51 example.

The only difference between f_1 and f_2, as well as f_3 and f_4, is which draw of a blue or green ball leads to a higher payoff of 200. Due to the 51st ball, when the likelihood of blue and green is assessed, f_2 and f_4 have a slight "objective advantage" compared to f_1 and f_3, respectively. Machina (2009) conjectures that plausible choices are $f_1 \succ f_2$ and $f_4 \succ f_3$. The first preference is motivated by the fact that f_2 is more ambiguous than f_1—ambiguity aversion offsets the objective advantage of f_2. The second might arise because both f_3 and f_4 are ambiguous—f_3 does not have an informational advantage as f_1 does. However, in most ambiguity-sensitive models, $f_1 \succ f_2$ implies $f_3 \succ f_4$ (Machina, 2009; Baillon et al., 2011). In Appendix C, we further extend the incomparability to a broader class of preferences.

We show that the costly ambiguity perception model can rationalize Machina's (2009) conjecture. Let $p = 50/101$, let $q = 1 - p$, let $B = [-p/2, p/2]$, and let $G = [-q/2, q/2]$. Identify each $(i, j) \in [0, p] \times [0, q]$ with the probability distribution over colors such that the probability of drawing blue is i and drawing green is j. For each $(\beta, \gamma) \in [0, 1]^2$, let $M(\beta, \gamma) = \{ (p/2 + \beta b, q/2 + \gamma g) \mid (b, g) \in B \times G \}$. Each of the parameters β and γ controls the size of ambiguity perception $M(\beta, \gamma)$. The bigger β is, the more ambiguous the belief about blue is. The same is true for γ. Let $\mathbb{M} = \{ M(\beta, \gamma) \mid (\beta, \gamma) \in [0, 1]^2 \}$. Define $c\colon [0, 1]^2 \to \mathbb{R}_+$ by $c(\beta, \gamma) = 25[2 - (\beta + \gamma)]$. Let U be the utility function over acts corresponding to the cost structure $(\mathbb{M}, c)$. Then,

$$
\begin{aligned}
U(f_1) &= 200p + 100q = 100(1 + p),\\
U(f_2) &= \max_{(\beta,\gamma)\in[0,1]^2} \Big[100 \min_{(b,g)\in B\times G} \Big(\frac{3}{2} - \beta b + \gamma g\Big) - c(\beta, \gamma)\Big]\\
&= \max_{(\beta,\gamma)\in[0,1]^2} [100 + (25 - 50p)\beta + (25 - 50q)\gamma] = 125 - 50p,\\
U(f_3) &= \max_{(\beta,\gamma)\in[0,1]^2} \Big[100 \min_{(b,g)\in B\times G} \Big(\frac{5}{2}p + \frac{1}{2}q - \beta r + \gamma g\Big) - c(\beta, \gamma)\Big]\\
&= \max_{(\beta,\gamma)\in[0,1]^2} [250p + 50q - 50 + (25 - 50p)\beta + (25 - 50q)\gamma] = 25 + 150p,\\
U(f_4) &= \max_{(\beta,\gamma)\in[0,1]^2} \Big[100 \min_{(b,g)\in B\times G} (2p + q - 2\beta r + 2\gamma g) - c(\beta, \gamma)\Big]\\
&= \max_{(\beta,\gamma)\in[0,1]^2} [200p + 100q - 50 - (25 - 100p)\beta + (25 - 100q)\gamma] = 50 + 100p.
\end{aligned}
$$

Thus, $f_1 \succ f_2$ and $f_4 \succ f_3$.

In this example, the optimal ambiguity perception at $f_1, \ldots, f_4$ are $M(1, 1)$, $M(1, 0)$, $M(1, 0)$, and $M(0, 0)$, respectively. The DM incurs no filtering cost to evaluate unambiguous act f_1, whereas she must bear a cost to filter her ambiguity perception optimally when evaluating at ambiguous act f_2. Here, the cost exceeds the objective advantage of f_2, so the DM prefers f_1 to f_2. In contrast, when f_4 is compared to f_3, the objective advantage of f_4 dominates the additional cost to filter the belief about blue. As a result, the DM prefers f_4 to f_3.

5.2. **The reflection example.** Machina's (2009) second thought experiment uses a box slightly modified from the previous one: 50 balls—instead of 51—are either green or purple. Table 2 describes acts $f_5, \ldots, f_8$.

The environment is informationally symmetric—there is no evidence to believe one color is more likely than another. Thus, f_7 is a "reflection" of f_6. The same is true for f_8 and f_5. In many models of ambiguity, to respect this informational symmetry, the DM has to be indifferent between all the four acts (Machina, 2009; Baillon et al., 2011). However, L'Haridon and Placido (2010) report that a typical pattern of an ambiguity-averse DM's choices are $f_6 \succ f_5$ and $f_7 \succ f_8$.

	50 balls		50 balls	
	Red	Blue	Green	Purple
f_5	100	200	100	0
f_6	100	100	200	0
f_7	0	200	100	100
f_8	0	100	200	100

TABLE 2. The reflection example.

To see that the costly ambiguity perception model can justify the typical pattern, let $B = G = [-1/4, 1/4]$. As in the previous example, identify each $(i, j) \in [0, 1/2] \times [0, 1/2]$ with the probability distribution over colors such that the probability of drawing blue is i and drawing green is j. For each $(\beta, \gamma) \in [0,1]^2$, let $M(\beta, \gamma) = \{ (1/4 + \beta b, 1/4 + \gamma g) \mid (b, g) \in B \times G \}$. Let $\mathbb{M} = \{ M(\beta, \gamma) \mid (\beta, \gamma) \in [0,1]^2 \}$. Define $c\colon [0,1]^2 \to \mathbb{R}_+$ by $c(\beta, \gamma) = 30[2 - (\beta + \gamma)]$. Let U be the utility function over acts corresponding to the cost structure $(\mathbb{M}, c)$. Then,

$$U(f_5) = \max_{(\beta,\gamma)\in[0,1]^2} \left[100 \min_{(b,g)\in B\times G} (1 + \beta b + \gamma g) - c(\beta, \gamma) \right] = \max_{(\beta,\gamma)\in[0,1]^2} (40 + 5\beta + 5\gamma) = 50,$$

$$U(f_6) = \max_{(\beta,\gamma)\in[0,1]^2} \left[100 \min_{(b,g)\in B\times G} (1 + 2\gamma g) - c(\beta, \gamma) \right] = \max_{(\beta,\gamma)\in[0,1]^2} (40 + 30\beta - 20\gamma) = 70,$$

$$U(f_7) = \max_{(\beta,\gamma)\in[0,1]^2} \left[100 \min_{(b,g)\in B\times G} (1 + 2\beta b) - c(\beta, \gamma) \right] = \max_{(\beta,\gamma)\in[0,1]^2} (40 - 20\beta + 30\gamma) = 70,$$

$$U(f_8) = \max_{(\beta,\gamma)\in[0,1]^2} \left[100 \min_{(b,g)\in B\times G} (1 + \beta b + \gamma g) - c(\beta, \gamma) \right] = \max_{(\beta,\gamma)\in[0,1]^2} (40 + 5\beta + 5\gamma) = 50.$$

Thus, $f_6 \succ f_5$ and $f_7 \succ f_8$.

The optimal ambiguity perception at $f_5, \ldots, f_8$ are $M(1,1)$, $M(1,0)$, $M(0,1)$, and $M(1,1)$, respectively. That is, the DM engages in costly filtering only when facing f_6 or f_7. The intuition is as follows. When facing f_5 or f_8, the DM has to be worried about the composition of both red and blue, and green and purple. In contrast, she can concentrate on filtering only about one side when facing f_6 or f_7. In this sense, f_5 (resp. f_8) is more “complex” than f_6 (resp. f_7). This additional complexity makes the evaluation of f_5 or f_8 lower.

The preference in this example can accommodate the typical Ellsberg pattern at the same time. Table 3 reproduces the two-color Ellsberg-type experiment in the current box. Act f_9 is risky while act f_{10} is ambiguous under informational symmetry. So it is expected that f_9 is preferred to f_{10}.

The utility of each act is given by

$$U(f_9) = 100 \times \frac{1}{2} = 50,$$

	50 balls		50 balls	
	Red	Blue	Green	Purple
f_9	100	100	0	0
f_{10}	0	100	100	0

TABLE 3. The Ellsberg-type example.

$$U(f_{10}) = \max_{(\beta,\gamma)\in[0,1]^2}\Big[100 \min_{(b,g)\in B\times G}(\beta b + \gamma g) - c(\beta,\gamma)\Big] = \max_{(\beta,\gamma)\in[0,1]^2}(5\beta + 5\gamma - 60) = -50,$$

which shows $f_9 \succ f_{10}$.

6. DISCUSSION

6.1. **Special cases.** We examine three special cases of the costly ambiguity perception model. Each of them has been studied under the simpler domain without ex ante randomization.

The first two cases restrict the DM's choice of perceptions.

A **convex capacity** is a supermodular real-valued function ν on 2^{Ω} such that $\nu(\emptyset) = 0$ and $\nu(\Omega) = 1$. For each convex capacity ν, the **core of** ν is the set $\{\, \mu \in \Delta(\Omega) \mid \mu(E) \geq \nu(E) \text{ for each } E \in 2^{\Omega} \,\}$, and is denoted by $\mathrm{core}(\nu)$. Every convex capacity has a nonempty core. An **optimal ambiguity perception representation of** $\succsim$ is a costly ambiguity perception representation $\langle u, (\mathbb{M}, c)\rangle$ where each member of $\mathbb{M}$ is the core of some convex capacity (Payró et al., 2025). For each $\phi \in \mathbb{R}^{\Omega}$, its Choquet integral $\int \phi \, \mathrm{d}\nu$ with respect to a convex capacity ν equals $\min_{\mu\in\mathrm{core}(\nu)} \int \phi \, \mathrm{d}\mu$. Thus, if $\succsim$ has an optimal ambiguity perception representation $\langle u, (\mathbb{M}, c)\rangle$, then there exists a pair (N, C) of a set of convex capacities and a grounded real-valued function on N such that

$$P \mapsto \max_{\nu\in N}\Big[\int \Big(\int u\circ f \, \mathrm{d}\nu\Big) \, \mathrm{d}P(f) - C(\nu)\Big]$$

represents $\succsim$.

The second case restricts possible ambiguity perceptions to singletons. A **moral-hazard representation of** $\succsim$ is a pair $\langle u, (M, C)\rangle$ of a surjective vNM function and a pair of a nonempty compact subset of $\Delta(\Omega)$ and a lower semicontinuous grounded real-valued function on M such that

$$P \mapsto \max_{\mu\in M}\Big[\int \Big(\int u\circ f \, \mathrm{d}\mu\Big) \, \mathrm{d}P(f) - C(\mu)\Big]$$

represents $\succsim$ (Sinander, 2024).

These two cases are characterized by strengthening *indifference to mixture timing of constant acts*. A pair $(f, g) \in \mathscr{F}^2$ is **comonotonic** if $\delta[f(\omega)] \succsim \delta[f(\omega')]$ is equivalent to $\delta[g(\omega)] \succsim \delta[g(\omega')]$ for each $(\omega, \omega') \in \Omega^2$.

Axiom 8 (Indifference to mixture timing of comonotonic acts)**.** For each $((\kappa,\lambda),R,(f,g)) \in [0,1]^2 \times \Delta_s(\mathscr{F}) \times \mathscr{F}^2$, if (f,g) is comonotonic, then $\kappa\delta[\lambda f + (1-\lambda)g] + (1-\kappa)R \sim \kappa[\lambda\delta[f] + (1-\lambda)\delta[g]] + (1-\kappa)R$.

Axiom 9 (Indifference to mixture timing)**.** For each $((\kappa,\lambda),R,(f,g)) \in [0,1]^2 \times \Delta_s(\mathscr{F}) \times \mathscr{F}^2$, we have $\kappa\delta[\lambda f + (1-\lambda)g] + (1-\kappa)R \sim \kappa[\lambda\delta[f] + (1-\lambda)\delta[g]] + (1-\kappa)R$.

In the third special case of the costly ambiguity perception model, the DM chooses her ambiguity perception without costs. A **dual-self expected utility representation of** $\succsim$ is a pair $(u,\mathbb{M})$ of a surjective vNM function and a nonempty compact subset of $\mathbb{K}$ such that

$$P \mapsto \max_{M\in\mathbb{M}} \int \left(\min_{\mu\in M} \int u \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f)$$

represents $\succsim$ (Chandrasekher et al., 2022).

The dual-self expected utility model is characterized by strengthening *independence of constant acts*.

Axiom 10 (Strong independence of constant acts)**.** For each $(\lambda,(P,Q),p) \in [0,1] \times \Delta_s(\mathscr{F})^2 \times \Delta_s(X)$, we have $P \succsim Q$ if and only if $\lambda P + (1-\lambda)\delta[p] \succsim \lambda Q + (1-\lambda)\delta[p]$.

The following corollary collects alternative characterizations of the three models on our domain.

Corollary 2. *Suppose that* $\succsim$ *has a costly ambiguity perception representation.*

(i) *The relation* $\succsim$ *satisfies indifference to mixture timing of comonotonic acts if and only if it has an optimal ambiguity perception representation.*

(ii) *The relation* $\succsim$ *satisfies indifference to mixture timing if and only if it has a moral-hazard representation.*

(iii) *The relation* $\succsim$ *satisfies strong independence of constant acts if and only if it has a dual-self expected utility representation.*

As mentioned in Payró et al. (2025), the optimal ambiguity perception model can explain the typical behaviors in the 50–51 and reflection examples. Indeed, the numerical examples in Section 5 fall within the class of optimal ambiguity perception models. Since every moral-hazard preference is ambiguity-seeking (Sinander, 2024), it is incompatible with the Ellsberg-type behavior. In Appendix C, we show that the dual-self expected utility model can accommodate the typical pattern in the reflection example, but fails to explain the one in the 50–51 example.

6.2. **Dual representation.** Our costly ambiguity perception model is closely related to the double maxmin expected utility model proposed by Ke and Zhang (2020); it is a “dual” version of the dual-self expected utility model. The domain of their model is the same as ours: the original Anscombe–Aumann domain with two-stage randomization.

A **cautious costly ambiguity perception representation of** $\succsim$ is a pair $\langle u, (\mathbb{M}, c)\rangle$ of a surjective vNM function on $\Delta_s(X)$ and a cost structure such that

$$P \mapsto \min_{M \in \mathbb{M}} \left[\int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f) + c(M) \right]$$

represents $\succsim$. A **double maxmin expected utility representation of** $\succsim$ is a pair $(u, \mathbb{M})$ of a surjective vNM function and a nonempty compact subset of $\mathbb{K}$ such that

$$P \mapsto \min_{M \in \mathbb{M}} \int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f)$$

represents $\succsim$. The former representation generalizes Ke and Zhang's (2020) latter representation by adding the cost term.

Ke and Zhang (2020) motivate the double maxmin expected utility model by considering the case where ex ante randomization may help the DM hedge against ambiguity because of the DM's subjective perception of the timing of randomization. The next is a key difference from the costly ambiguity perception model.

Axiom 11 (Ex ante attraction to randomization)**.** For each $(\lambda, (P, Q)) \in [0, 1] \times \Delta_s(\mathscr{F})^2$, if $P \succsim Q$, then $\lambda P + (1 - \lambda)Q \succsim Q$.

By adopting *ex ante attraction to randomization* instead of *ex ante aversion to randomization*, we can replace the "max" in the costly ambiguity perception representation with "min".

Corollary 3. *Assume regularity, FSD, attraction to ex post randomization, indifference to mixture timing of constant acts, independence of constant acts, and unboundedness.*

(i) *The relation $\succsim$ satisfies ex ante attraction to randomization if and only if it has a cautious costly ambiguity perception representation.*

(ii) *The relation $\succsim$ satisfies ex ante attraction to randomization and strong independence of constant acts if and only if it has a double maxmin expected utility representation.*

We omit the proof of Corollary 3 since the same argument as in the proof of Theorem 1 and Corollary 2 (iii) applies by changing "convexity" to "concavity" and "max" to "min".

In addition to the set of axioms in Corollary 3, to characterize the double maxmin expected utility model, Ke and Zhang (2020) impose the following axiom.

Axiom 12 (Preference for statewise randomization)**.** For each $(\lambda, (f, g)) \in (0, 1) \times \mathscr{F}^2$,

(i) $\delta[f] \succsim \delta[g]$ implies $\delta[\lambda f + (1 - \lambda)g] \succsim \delta[g]$;

(ii) $\delta[f] \succsim \delta[g]$ if and only if $\delta[\lambda f + (1 - \lambda)p] \succsim \delta[\lambda g + (1 - \lambda)p]$ for each $p \in \Delta_s(X)$.

That is, the restriction of $\succsim$ to $\mathscr{F}$ is required to satisfy the *uncertainty aversion* and *certainty independence* axioms. As we can see in Corollary 3 (ii), it turns out that *preference for statewise randomization* is redundant under the other axioms. We can check this point in the level of axioms.

Proposition 5. *If $\succsim$ satisfies regularity, attraction to ex post randomization, indifference to mixture timing of constant acts, ex ante attraction to randomization, and strong independence of constant acts, then it satisfies preference for statewise randomization.*

By a similar reasoning, we can see that the cautious costly ambiguity perception preference satisfies the *weak certainty independence* axiom of Maccheroni et al. (2006). Thus, it coincides with the variational model (Maccheroni et al., 2006) on the domain of acts, and so it cannot account for Machina's examples, as shown in Baillon et al. (2011).

Appendix A. Preliminaries

A.1. **Convex analysis.** Let E be a normed space, and let E' be its norm dual. Define the bilinear functional $\langle\cdot,\cdot\rangle$ on $E\times E'$ by $\langle x,x'\rangle = x'(x)$.

For each real-valued function W on a subset D of E, the **convex conjugate of** W is the extended real-valued function W^* on E' of the form $W^*(x') = \sup_{x\in D}(\langle x,x'\rangle - W(x))$; the **subdifferential of** W is the correspondence $\partial W\colon D \rightrightarrows E'$ of the form $\partial W(x) = \bigcap_{y\in D}\{\, x'\in E' \mid W(y)\geq W(x) + \langle y-x,x'\rangle\,\}$; let $\operatorname{gr}\partial W$ be the graph $\{\,(x,x')\in D\times E' \mid x'\in\partial W(x)\,\}$ of ∂W.

Lemma 1. *Let W be a real-valued function on a subset D of E.*

(i) *W^* is weak$*$ lower semicontinuous and convex.*

(ii) *$W(x)\geq\langle x,x'\rangle - W^*(x')$ for each $(x,x')\in D\times E'$.*

(iii) *$W(x)=\langle x,x'\rangle - W^*(x')$ if and only if $(x,x')\in\operatorname{gr}\partial W$.*

(iv) *∂W has weak$*$ closed convex values.*

(v) *If W is lower semicontinuous and if B is a norm bounded weak$*$ closed subset of E', then $\operatorname{gr}\partial W\cap(D\times B)$ is norm $\times$ weak$*$ closed.*

Proof. (i) Since W^* is the pointwise supremum of the family $(\langle x,\cdot\rangle - W(x))_{x\in E}$ of weak$*$ continuous affine functions, it is weak$*$ lower semicontinuous and convex.

(ii) For each $(x,x')\in D\times E'$, since $W^*(x')\geq\langle x,x'\rangle - W(x)$, we have $W(x)\geq\langle x,x'\rangle - W^*(x')$.

(iii) It follows that $(x,x')\in\operatorname{gr}\partial W$ if and only if $W(x)\leq\langle x,x'\rangle - (\langle y,x'\rangle - W(y))$ for each $y\in D$, which is equivalent to $W(x)\leq\langle x,x'\rangle - W^*(x')$. The statement follows from part (ii).

(iv) For each $x\in D$, the set $\partial W(x)$ is the intersection of closed half spaces.

(v) By Theorem 2.4.2 (ix) of Zălinescu (2002). □

A convex function W is **proper** if $\operatorname{dom} W\neq\emptyset$ and never assumes the value $-\infty$.

Lemma 2. *Let W be a Lipschitz continuous proper convex function on a convex subset D of E.*

(i) *∂W has nonempty values.*

(ii) *For each $(x, y) \in D^2$, it follows that $\partial W(x) \cap \partial W(y) \neq \emptyset$ if and only if $W(\lambda x + (1-\lambda)y) = \lambda W(x) + (1-\lambda)W(y)$ for each $\lambda \in [0,1]$.*

PROOF. (i) Apply the same argument as in the proof of the Duality Theorem of Gale (1967).
(ii) Apply the same argument as in the proof of Proposition 1 of Pennesi (2015). □

A subset C of E is **conical** if $\alpha C \subseteq C$ for each $\alpha \in \mathbb{R}_+$. A **cone** is a conical subset of E. For each subset S of E, let $S^+ = \bigcap_{x \in S}\{ x' \in E' \mid \langle x, x'\rangle \geq 0 \}$, and let $S^{++} = \bigcap_{x' \in S^+}\{ x \in E \mid \langle x, x'\rangle \geq 0 \}$. The next is Theorem 1.1.9 of Zălinescu (2002).

Lemma 3. *The closure of a convex conical subset S of E equals S^{++}.*

Let W be a convex function on a convex subset D of E. For each $(x, v) \in D \times E$ with $x + \varepsilon v \in D$ for some $\varepsilon \in \mathbb{R}_{++}$, the **right directional derivative of W in the direction v at x** is the number $\mathrm{D}_v^+ W(x)$ defined as

$$\mathrm{D}_v^+ W(x) = \lim_{\lambda \downarrow 0} \frac{W(x + \lambda v) - W(x)}{\lambda}.$$

The following comes from Theorem 2.4.9 of Zălinescu (2002).

Lemma 4. *Let W be a continuous proper convex function on E. Then, $\mathrm{D}_v^+ W(x) = \max_{x' \in \partial W(x)} \langle v, x'\rangle$ for each $(x, v) \in E^2$.*

For each real-valued function W on a convex subset D of E, define the relation $\succcurlyeq_W^*$ on D by

$$x \succcurlyeq_W^* y \iff W(\lambda x + (1-\lambda)z) \geq W(\lambda y + (1-\lambda)z) \quad \forall (\lambda, z) \in (0,1] \times D,$$

whose graph $\{ (x, y) \in D^2 \mid x \succcurlyeq_W^* y \}$ is denoted by $\mathrm{gr}(\succcurlyeq_W^*)$. A binary relation $\succcurlyeq$ on D is **mixture independent** if $x \succcurlyeq \tilde{x}$ if and only if $\lambda x + (1-\lambda)y \succcurlyeq \lambda \tilde{x} + (1-\lambda)y$ for each $(\lambda, y) \in (0,1] \times D$.

Lemma 5. *Let W be a real-valued function on a convex subset D of E. Then, $\succcurlyeq_W^*$ is transitive and mixture independent.*

PROOF. By definition, $\succcurlyeq_W^*$ is transitive. If $x \succcurlyeq_W^* \tilde{x}$, then for each $(\lambda, y) \in (0,1] \times D$, since for each $(\kappa, z) \in (0,1] \times D$, letting $\tilde{z} = (1-\kappa\lambda)^{-1}[\kappa(1-\lambda)y + (1-\kappa)z]$ gives $\tilde{z} \in D$ and

$$\begin{aligned} W(\kappa[\lambda x + (1-\lambda)y] + (1-\kappa)z) &= W(\kappa\lambda x + (1-\kappa\lambda)\tilde{z}) \\ &\geq W(\kappa\lambda\tilde{x} + (1-\kappa\lambda)\tilde{z}) = W(\kappa[\lambda\tilde{x} + (1-\lambda)y] + (1-\kappa)z), \end{aligned}$$

we have $\lambda x + (1-\lambda)y \succcurlyeq_W^* \lambda\tilde{x} + (1-\lambda)y$. Thus, $\succcurlyeq_W^*$ is mixture independent. □

A.2. **Niveloids.** Let $(E, \geq)$ be a Riesz space with unit e; that is, E is a lattice under the order $\geq$, and for each $x \in E$, there exists $\alpha \in \mathbb{R}_{++}$ such that $\alpha e \geq x \geq -\alpha e$. Denote by E_+ the positive cone of E. For each $x \in E$, the **absolute value of** x is $x \vee (-x)$, and is denoted by $|x|$; the **essential supremum of** x is the number $\inf\{\, \alpha \in \mathbb{R} \mid \alpha e \geq x \,\}$, and is denoted by $\operatorname{esup}(x)$. Endow E with the norm $x \mapsto \operatorname{esup}(|x|)$.

A real-valued function W on a subset D of E is

- **monotone** if $W(x) \geq W(y)$ for each $(x, y) \in D^2$ with $x \geq y$;
- **normalized** if $W(te) = t$ for each $t \in \mathbb{R}$ with $te \in D$;
- **translation equivariant** if $W(x + te) = W(x) + t$ for each $(x, t) \in D \times \mathbb{R}$ with $x + te \in D$;
- a **niveloid** if $W(x) - W(y) \leq \operatorname{esup}(x - y)$ for each $(x, y) \in D^2$.

By construction, every niveloid is 1-Lipschitz continuous.

A **tube** is a subset T of E such that $T + \{\, te \mid t \in \mathbb{R} \,\} \subseteq T$. The next is Proposition 2 of Cerreia-Vioglio et al. (2014).

Lemma 6.

(i) *Every niveloid is monotone and translation equivariant.*

(ii) *Every monotone translation equivariant real-valued function on a tube is a niveloid.*

Let $\Delta = \{\, x' \in E'_+ \mid \langle e, x' \rangle = 1 \,\}$, which is weak$*$ compact and convex. For each real-valued function W on a subset D of E, define the correspondence $\partial_\pi W \colon D \rightrightarrows E$ by $\partial_\pi W(x) = \partial W(x) \cap \Delta$.

Lemma 7. *Let W be a convex niveloid on a convex subset D of E.*

(i) *If $D = E$, then $\partial W = \partial_\pi W$.*

(ii) *$\partial_\pi W$ has nonempty weak$*$ compact convex values.*

Proof. (i) Suppose $D = E$. By Lemma 6 (i), W is monotone and translation equivariant. Thus, for each $(x, x') \in \operatorname{gr} \partial W$, since $\langle -v, x' \rangle = \langle x - v, x' \rangle - \langle x, x' \rangle \leq W(x - v) - W(x) \leq 0$ for each $v \in E_+$, we have $x' \in E'_+$; since $\alpha \langle e, x' \rangle = \langle x + \alpha e, x' \rangle - \langle x, x' \rangle \leq W(x + \alpha e) - W(x) = \alpha$ for each $\alpha \in \mathbb{R}$, we have $\langle e, x' \rangle = 1$.

(ii) By Theorem 1 and Proposition 4 of Cerreia-Vioglio et al. (2014), W has a convex niveloidal extension $\hat{W}$ to E. For each $x \in D$, since $\partial W(x)$ is weak$*$ closed and convex by Lemma 1 (iv), $\partial_\pi W(x)$ is weak$*$ compact and convex; since $\partial \hat{W}(x) \neq \emptyset$ by Lemma 2 (i) and since $\partial_\pi \hat{W}(x) = \partial \hat{W}(x) \subseteq \partial W(x)$ by part (i), we have $\partial_\pi W(x) \neq \emptyset$. □

For each real-valued function W on a subset D of E, a **variational representation of** W is a pair (Π, γ) of a weak$*$ compact subset of Δ and a weak$*$ lower semicontinuous extended real-valued function on Π such that $W(x) = \max_{\pi \in \Pi}(\langle x, \pi \rangle - \gamma(\pi))$ for each $x \in D$.

Lemma 8. *Let W be a real-valued function on a convex subset D of E.*

(i) *If W has a variational representation, then it is convex, monotone, and translation equivariant.*

(ii) *If $te \in D$ for some $t \in \mathbb{R}$, if W is normalized, and if (Π, γ) is a variational representation of W, then γ is grounded.*

Proof. (i) If (Π, γ) is a variational representation of W, then since $\langle \cdot, \pi \rangle - \gamma(\pi)$ is mixture linear, monotone, and translation equivariant for each $\pi \in \Pi$, the function W is convex, monotone, and translation equivariant.

(ii) Suppose that $te \in D$ for some $t \in \mathbb{R}$, that W is normalized, and that (Π, γ) is a variational representation of W. Since $t = W(te) = \max_{\pi \in \Pi}(\langle te, \pi \rangle - \gamma(\pi)) = t - \min \gamma(\Pi)$, the function γ is grounded. □

For each real-valued function W on a convex subset of E, a **multi-expectation representation of $\succcurlyeq^*_W$** is a subset Π of Δ such that

$$x \succcurlyeq^*_W y \iff \langle x, \pi \rangle \geq \langle y, \pi \rangle \quad \forall \pi \in \Pi.$$

Proposition 6. *Let W be a convex niveloid on a convex subset D of E. Then, there exists a weak∗ compact convex subset Π of* $\operatorname{dom} W^*|_\Delta$ *such that Π is a multi-expectation representation of $\succcurlyeq^*_W$ and $(\Pi, W^*|_\Pi)$ is a variational representation of W.*

Proof. Let $K = \{ \alpha(x-y) \mid (\alpha, (x,y)) \in \mathbb{R}_+ \times \operatorname{gr}(\succcurlyeq^*_W) \}$. By construction, K is conical. For each $((\alpha, (x,y)), (\beta, (\tilde{x}, \tilde{y}))) \in (\mathbb{R}_{++} \times \operatorname{gr}(\succcurlyeq^*_W))^2$, letting $\lambda = \alpha/(\alpha+\beta)$ gives $\lambda x + (1-\lambda) y \succcurlyeq^*_W \lambda y + (1-\lambda)\tilde{x} \succcurlyeq^*_W \lambda y + (1-\lambda)\tilde{y}$ by Lemma 5, so $\alpha(x-y) + \beta(\tilde{x} - \tilde{y}) = (\alpha+\beta)[\lambda(x-y) + (1-\lambda)(\tilde{x} - \tilde{y})] \in K$. Thus, K is convex. Since W is continuous, the set K is closed. Thus, by Lemma 3, $K = K^{++}$.

We claim that $\partial_\pi W(x)$ intersects K^+ for each $x \in D$. Seeking a contradiction, suppose otherwise. Let $x \in D$ be such that $\partial_\pi W(x) \cap K^+ = \emptyset$. By Lemma 7 (ii), we can apply the separation theorem (Aliprantis and Border, 2006, Theorem 5.79) to get $(\bar{x}, c) \in E \times \mathbb{R}$ such that $\bar{x} \neq 0$ and $\langle \bar{x}, \pi \rangle < c \leq \langle \bar{x}, x' \rangle$ for each $(\pi, x') \in \partial_\pi W(x) \times K^+$. Since K^+ is conical, we have $\bar{x} \in K^{++} = K$ and $c \leq 0$. Since by Lemma 1 (v) and Theorem 17.11 of Aliprantis and Border (2006), the correspondence $\partial_\pi W$ is upper hemicontinuous, there exists an open neighborhood U of x such that $\partial_\pi W(u) \subseteq \{ \pi \in \Delta \mid \langle \bar{x}, \pi \rangle < c \}$ for each $u \in U$. Let $(u, \varepsilon) \in U \times \mathbb{R}_{++}$ be such that $u - \varepsilon\bar{x} \in D$. Since $\bar{x} \in K$, we have $u \succcurlyeq^*_W u - \varepsilon\bar{x}$. However, $W(u - \varepsilon\bar{x}) - W(u) \geq -\langle \varepsilon\bar{x}, \pi \rangle > -\varepsilon c \geq 0$ for each $\pi \in \partial_\pi W(u)$, which is a contradiction.

Let $\Pi = \operatorname{dom} W^*|_\Delta \cap K^+$. By the above claim and Lemma 1 (i)–(iii), Π is nonempty, weak∗ compact, and convex, and $(\Pi, W^*|_\Pi)$ is a variational representation of W. Thus, for each $(x,y) \in D^2$, if $\langle x, \pi \rangle \geq \langle y, \pi \rangle$ for each $\pi \in \Pi$, then $x \succcurlyeq^*_W y$. Since $x \succcurlyeq^*_W y$ implies $\langle x, x' \rangle \geq \langle y, x' \rangle$ for each $x' \in K^+$, the set Π is a multi-expectation representation of $\succcurlyeq^*_W$. □

A variational representation (Π, γ) is **canonical** if

(i) for each $(\pi, \pi') \in \Pi^2$, if $\langle x, \pi \rangle \geq \langle x, \pi' \rangle$ for each $x \in D$, then $\gamma(\pi) \geq \gamma(\pi')$;

(ii) Π and γ are convex.

The next generalizes Theorem 2 of de Oliveira et al. (2017) to an abstract setting.

Proposition 7. *Let (Π, γ) be a canonical variational representation of a convex niveloid W on a convex conical subset D in E. Then, $\gamma = W^*|_{\Pi}$.*

PROOF. For each $\pi \in \Pi$, since $\gamma(\pi) \geq \langle x, \pi\rangle - W(x)$ for each $x \in D$, we have $\gamma(\pi) \geq W^*(\pi)$. Thus, $\gamma \geq W^*|_{\Pi}$. For the converse inequality, fix any $\bar{\pi} \in \Pi$. Choose any $\alpha \in [\min W^*(\Pi), \gamma(\bar{\pi}))$. Define $\hat{W} : D^{++} \to \mathbb{R}$ by $\hat{W}(x) = \max_{\pi\in\Pi}(\langle x, \pi\rangle - \gamma(\pi))$, which is a unique niveloidal extension of W to D^{++}. By the definition of W^*, it suffices to show that there exists $\bar{x} \in D^{++}$ for which $\langle \bar{x}, \bar{\pi}\rangle - \hat{W}(\bar{x}) > \alpha$. Let $C = \{ (\bar{\pi} + x', \alpha) \mid x' \in D^+ \}$, let $\operatorname{epi}\gamma$ be the epigraph of γ, and let $F = \operatorname{epi}\gamma \cap (\Pi\times[\min W^*(\Pi), \alpha+1]) - C$. For each $x' \in D^+$ with $\bar{\pi}+x' \in \Pi$, since $\langle x, \bar{\pi}+x'\rangle \geq \langle x, \bar{\pi}\rangle$ for each $x \in D$, we have $\gamma(\bar{\pi} + x') \geq \gamma(\bar{\pi}) > \alpha$. Thus, $\operatorname{epi}\gamma \cap C = \emptyset$, so F does not contain zero. Since $\operatorname{epi}\gamma \cap (\Pi \times [\alpha + 1, \infty)) - C \subseteq E' \times [1, \infty)$, we have $\operatorname{epi}\gamma - C \subseteq F \cup (E' \times [1, \infty))$. Hence, the weak$*$ closure of $\operatorname{epi}\gamma - C$ does not contain zero. Therefore, we can apply the separation theorem (Aliprantis and Border, 2006, Corollary 5.80) to get $(\bar{x}, \beta) \in E \times \mathbb{R}$ such that

$$\sup_{((\pi,r),x')\in\operatorname{epi}\gamma\times D^+} (\langle \bar{x}, \pi - \bar{\pi} - x'\rangle + \beta(r - \alpha)) < 0,$$

which implies $(\bar{x}, \beta) \in D^{++} \times (-\mathbb{R}_+)$, so

$$\max_{\pi\in\Pi}(\langle \bar{x}, \pi\rangle + \beta\gamma(\pi)) < \langle \bar{x}, \bar{\pi}\rangle + \alpha\beta.$$

If $\beta = 0$, then $\max_{\pi\in\Pi}\langle \bar{x}, \pi\rangle < \langle \bar{x}, \bar{\pi}\rangle$, so for sufficiently large $n \in \mathbb{N}$,

$$\langle n\bar{x}, \bar{\pi}\rangle - \hat{W}(n\bar{x}) = \langle n\bar{x}, \bar{\pi}\rangle - \max_{\pi\in\Pi}(\langle n\bar{x}, \pi\rangle - \gamma(\pi)) \geq n\Big(\langle \bar{x}, \bar{\pi}\rangle - \max_{\pi\in\Pi}\langle \bar{x}, \pi\rangle\Big) + \min\gamma(\Pi) > \alpha.$$

If $\beta < 0$, then since $-\beta^{-1}\bar{x} \in D^{++}$, it is without loss of generality to assume $\beta = -1$, so $\alpha < \langle \bar{x}, \bar{\pi}\rangle - \max_{\pi\in\Pi}(\langle \bar{x}, \pi\rangle - \gamma(\pi)) = \langle \bar{x}, \bar{\pi}\rangle - \hat{W}(\bar{x})$. □

Lemma 9. *Let W be a convex niveloid on a convex conical subset D of E, and let Π be a subset of Δ such that $(\Pi, W^*|_{\Pi})$ is a canonical variational representation of W.*

(i) *For each $(x, y) \in D^2$, it follows that $\partial W(x)\cap\partial W(y)\cap\Pi \neq \emptyset$ if and only if $W(\lambda x+(1-\lambda)y) = \lambda W(x) + (1 - \lambda)W(y)$ for each $\lambda \in [0, 1]$.*

(ii) *$\mathrm{D}^+_v W(x) = \max_{\pi\in\partial W(x)\cap\Pi}\langle v, \pi\rangle$ for each $(x, v) \in D^2$.*

PROOF. Define $\hat{W} : E \to \mathbb{R}$ by $\hat{W}(x) = \max_{\pi\in\Pi}(\langle x, \pi\rangle - W^*(\pi))$, which is a convex niveloidal extension of W by Lemma 6 (ii) and Lemma 8 (i). Define $\gamma : \Delta \to (-\infty, \infty]$ by $\gamma(\pi) = W^*(\pi)$ if $\pi \in \Pi$ and $\gamma(\pi) = \infty$ otherwise. Then, (Δ, γ) is a canonical variational representation of $\hat{W}$, so by Proposition 7, $\gamma = \hat{W}^*|_{\Delta}$. Hence, by Lemma 1 (iii), $\partial\hat{W}(x) \subseteq \Pi$ for each $x \in E$.

(i) Choose any $(x, y) \in D^2$. By Lemma 2 (ii), $\partial W(x) \cap \partial W(y) \cap \Pi \neq \emptyset$ implies $W(\lambda x + (1 - \lambda)y) = \lambda W(x) + (1 - \lambda)W(y)$ for each $\lambda \in [0, 1]$. Conversely, suppose that $W(\lambda x + (1 - \lambda)y) =$

$\lambda W(x) + (1-\lambda)W(y)$ for each $\lambda \in [0,1]$. Then, $\hat{W}(\lambda x + (1-\lambda)y) = \lambda\hat{W}(x) + (1-\lambda)\hat{W}(y)$ for each $\lambda \in [0,1]$. Thus, again by Lemma 2 (ii), $\partial\hat{W}(x) \cap \partial\hat{W}(y) \neq \emptyset$. Here, by construction, $\partial W(x) \cap \partial W(y) \cap \Pi \supseteq \partial\hat{W}(x) \cap \partial\hat{W}(y)$.

(ii) Choose any $(x, v) \in D^2$. Since $\partial W(x) \supseteq \partial\hat{W}(x)$ by definition, it follows from Lemma 4 that $\mathrm{D}_v^+ W(x) = \mathrm{D}_v^+ \hat{W}(x) = \max_{\pi \in \partial\hat{W}(x)} \langle v, \pi\rangle \leq \max_{\pi \in \partial W(x) \cap \Pi} \langle v, \pi\rangle$. For each $\pi \in \partial W(x) \cap \Pi$, since $\lambda\langle v, \pi\rangle \leq W(x + \lambda v) - W(x)$ for each $\lambda \in \mathbb{R}_{++}$, we have $\langle v, \pi\rangle \leq \mathrm{D}_v^+ W(x)$. Thus, $\mathrm{D}_v^+ W(x) = \max_{\pi \in \partial W(x) \cap \Pi} \langle v, \pi\rangle$. □

A.3. **Cost structures.** Given any topological space Y, denote by $\mathrm{C_b}(Y)$ the Banach lattice of all bounded continuous real-valued functions on Y, endowed with the pointwise ordering and the uniform norm; denote by $ca(Y)$ the Banach lattice of all signed Borel measures on Y of bounded variation, endowed with the setwise ordering and the total variation norm; by the Riesz representation theorem (Aliprantis and Border, 2006, Corollary 14.15), $ca(Y)$ is the norm dual of $\mathrm{C_b}(Y)$ when Y is compact and metrizable.

Let $\Phi = \mathbb{R}^\Omega$, endowed with the uniform norm. The constant function $\mathbf{1}_\Omega$ of value one is an order unit of Φ. A real-valued function V on Φ is **positively homogeneous** if $V(\alpha x) = \alpha V(x)$ for each $(\alpha, x) \in \mathbb{R}_+ \times \Phi$.

Fix any $\bar{\omega} \in \Omega$. Let $\Psi = \{\, \psi \in \Phi \mid \psi(\bar{\omega}) = 0 \text{ and } \max_{\omega \in \Omega}|\psi(\omega)| = 1 \,\}$, and let $\mathbb{U}$ be the closed unit ball in the set of continuous real-valued functions on Ψ endowed with the uniform norm. The constant function $\mathbf{1}_\mathbb{U}$ of value one is an order unit of $\mathrm{C_b}(\mathbb{U})$, and the norm of each $\xi \in \mathrm{C_b}(\mathbb{U})$ equals $\mathrm{esup}(|\xi|)$. For each $\phi \in \Phi \setminus \{\, t\mathbf{1}_\Omega \mid t \in \mathbb{R} \,\}$, there exists a unique $(\alpha, \psi, t) \in \mathbb{R}_{++} \times \Psi \times \mathbb{R}$ such that $\phi = \alpha\psi + t\mathbf{1}_\Omega$. Thus, every $v \in \mathbb{U}$ has a unique positively homogeneous translation equivariant extension $\hat{v}$ to Φ, which has the form $\hat{v}(\alpha\psi + t\mathbf{1}_\Omega) = \alpha v(\psi) + t$. For each $m \in \Delta_\mathrm{s}(\Phi)$, define $m^\vee \in \mathrm{C_b}(\mathbb{U})$ by $m^\vee(v) = \int \hat{v}\, \mathrm{d}m$. Let $\Xi = \{\, m^\vee \mid m \in \Delta_\mathrm{s}(\Phi) \,\}$.

Lemma 10.

(i) *$m \mapsto m^\vee$ from $\Delta_\mathrm{s}(\Phi)$ to $\mathrm{C_b}(\mathbb{U})$ is mixture linear.*

(ii) *$\phi \mapsto \delta[\phi]^\vee$ from Φ to $\mathrm{C_b}(\mathbb{U})$ is positively homogeneous and translation equivariant.*

(iii) *Ξ is a convex conical tube in $\mathrm{C_b}(\mathbb{U})$.*

Proof. (i) For each $(\lambda, (l, m), v) \in [0,1] \times \Delta_\mathrm{s}(\Phi)^2 \times \mathbb{U}$, since $[\lambda l^\vee + (1-\lambda)m^\vee](v) = \lambda \int \hat{v}\, \mathrm{d}l + (1-\lambda)\int \hat{v}\, \mathrm{d}m = \int \hat{v}\, \mathrm{d}[\lambda l + (1-\lambda)m] = [\lambda l + (1-\lambda)m]^\vee(v)$.

(ii) By the positive homogeneity and translation equivariance of $\hat{v}$ for each $v \in \mathbb{U}$.

(iii) By part (i), the set Ξ is convex; since for each $(\alpha, m) \in \mathbb{R}_{++} \times \Delta_\mathrm{s}(\Phi)$, letting $l(F) = m(\alpha^{-1}F)$ for each $F \subseteq \Phi$ gives $\alpha m^\vee = l^\vee$, the set Ξ is conical; since for each $(\alpha, m) \in \mathbb{R}_{++} \times \Delta_\mathrm{s}(\Phi)$, letting $l(F) = m(F - \{t\mathbf{1}_\Omega\})$ for each $F \subseteq \Phi$ gives $m^\vee + t\mathbf{1}_\mathbb{U} = l^\vee$, the set Ξ is a tube. □

A real-valued function on Φ is **superlinear** if it is positively homogeneous and concave. Let $\mathbb{V}$ be the set of all superlinear niveloids on Φ. Endow $\mathbb{V}$ with the metric ρ of the form

$$\rho(V, V') = \max_{\phi \in \mathbb{B}} |V(\phi) - V'(\phi)|,$$

where $\mathbb{B}$ is the closed unit ball of Φ.

Lemma 11. *The function $V \mapsto V|_\Psi$ is a mixture linear isometry from $\mathbb{V}$ to $\mathbb{U}$.*

Proof. The mixture linearity holds by construction. For each $V \in \mathbb{V}$, since $V|_\Psi$ is continuous and since $1 = V(\mathbf{1}_\Omega) \geq V(\psi)$ for each $\psi \in \Psi$, we have $V|_\Psi \in \mathbb{U}$. Let $B = \{ (\alpha, \psi, t) \in \mathbb{R}_+ \times \Psi \times \mathbb{R} \mid \alpha\psi + t\mathbf{1}_\Omega \in \mathbb{B} \}$. For each $(V, V') \in \mathbb{V}^2$,

$$\begin{aligned} \rho(V, V') &= \max_{(\alpha,\psi,t)\in B} |V(\alpha\psi + t\mathbf{1}_\Omega) - V'(\alpha\psi + t\mathbf{1}_\Omega)| = \max_{(\alpha,\psi,t)\in B} \alpha|V(\psi) - V'(\psi)| \\ &= \max_{\psi\in\Psi} |V(\psi) - V'(\psi)|, \end{aligned}$$

so $V \mapsto V|_\Psi$ is an isometry. □

For each $M \in \mathbb{K}$, the **support function of** M is the real-valued function H_M on Φ of the form $H_M(\phi) = \min_{\mu\in M} \langle \phi, \mu \rangle$.

Lemma 12.

(i) *$M \mapsto H_M$ on $\mathbb{K}$ is a surjective mixture linear isometry to $\mathbb{V}$.*

(ii) *For each $(M, M') \in \mathbb{K}^2$, it follows that $M \subseteq M'$ if and only if $H_M \geq H_{M'}$.*

Proof. (i) For each $M \in \mathbb{K}$, since H_M is superlinear by Theorem 7.51 of Aliprantis and Border (2006) and since H_M is monotone and translation equivariant by the monotonicity and translation equivariance of $\langle \cdot, \mu \rangle$ for each $\mu \in M$, we have $H_M \in \mathbb{V}$. Since for each $(\lambda, (M, M'), \phi) \in [0, 1] \times \mathbb{K}^2 \times \Phi$,

$$H_{\lambda M+(1-\lambda)M'}(\phi) = \min_{(\mu,\mu')\in M\times M'} [\lambda\langle \phi, \mu \rangle + (1 - \lambda)\langle \phi, \mu' \rangle] = \lambda H_M(\phi) + (1 - \lambda)H_{M'}(\phi),$$

the function $M \mapsto H_M$ is mixture linear; it is surjective by Theorem 7.52 of Aliprantis and Border (2006) and Lemma 7 (i); it is an isometry by Corollary 7.59 of Aliprantis and Border (2006).

(ii) By Theorem 2.4.14 (vi) of Zălinescu (2002). □

Recall $\Delta = \{ \pi \in \mathrm{C_b}(\mathbb{U})'_+ \mid \langle \mathbf{1}_\mathbb{U}, \pi \rangle = 1 \}$. For each $\pi \in \Delta$, define $\pi^\wedge : \Phi \to \mathbb{R}$ by $\pi^\wedge(\phi) = \langle \delta[\phi]^\vee, \pi \rangle$. By definition, for each $(M, \pi) \in \mathbb{K} \times \Delta$ with $H_M = \pi^\wedge$, it follows that $\int H_M \, \mathrm{d}m = \int \pi^\wedge \, \mathrm{d}m = \langle m^\vee, \pi \rangle$ for each $m \in \Delta_\mathrm{s}(\Phi)$. For each subset Π of Δ, let $\Pi^\wedge = \{ \pi^\wedge \mid \pi \in \Pi \}$. Let $\Delta_\mathbb{V} = \{ \pi \in \Delta \mid \pi^\wedge \in \mathbb{V} \}$, endowed with the weak∗ topology.

Lemma 13. *The function $\pi \mapsto \pi^\wedge$ from $\Delta_\mathbb{V}$ to $\mathbb{V}$ is surjective, mixture linear, and continuous.*

PROOF. The mixture linearity follows by definition. Since $\mathrm{C_b}(\mathbb{U})'$ can be identified with the space of normal signed charges on the Borel algebra on $\mathbb{U}$ by the Riesz representation theorem (Aliprantis and Border, 2006, Theorem 14.10) and since $\delta[V|_{\Psi}]^{\wedge} = V$ for each $V \in \mathbb{V}$, the function $\pi \mapsto \pi^{\wedge}$ is surjective. For the continuity, choose any net $(\pi_d)_{d\in\mathbb{D}}$ in $\Delta_{\mathbb{V}}$ with a limit $\bar{\pi} \in \Delta_{\mathbb{V}}$. We show that $(\rho(\pi_d^{\wedge}, \bar{\pi}^{\wedge}))_{d\in\mathbb{D}}$ converges to 0. Fix any $\varepsilon > 0$. Let F be a finite subset of $\mathbb{B}$ whose $(\varepsilon/3)$-neighborhood includes $\mathbb{B}$. By the 1-Lipschitz continuity of each member of $\mathbb{V}$, for each $(\phi, d) \in \mathbb{B} \times \mathbb{D}$, there exists $\psi \in F$ such that

$$\begin{aligned}|\pi_d^{\wedge}(\phi) - \bar{\pi}^{\wedge}(\phi)| &\le |\pi_d^{\wedge}(\phi) - \pi_d^{\wedge}(\psi)| + |\pi_d^{\wedge}(\psi) - \bar{\pi}^{\wedge}(\psi)| + |\bar{\pi}^{\wedge}(\psi) - \bar{\pi}^{\wedge}(\phi)| \\ &< |\langle \delta[\psi]^{\vee}, \pi_d - \bar{\pi}\rangle| + \frac{2}{3}\varepsilon.\end{aligned}$$

Thus, since $(\pi_d)_{d\in\mathbb{D}}$ is eventually in the weak$*$ neighborhood $\bigcap_{\psi\in F}\{\,\tilde{\pi} \in \Delta_{\mathbb{V}} \mid |\langle \delta[\psi]^{\vee}, \tilde{\pi} - \bar{\pi}\rangle| < \varepsilon/3\,\}$ of $\bar{\pi}$, the net $(\rho(\pi_d^{\wedge}, \bar{\pi}^{\wedge}))_{d\in\mathbb{D}}$ is eventually in $[-\varepsilon, \varepsilon]$. □

For each cost structure $(\mathbb{M}, c)$, let $\mathbb{M}^{\diamond} = \{\,\pi \in \Delta \mid \pi^{\wedge} = H_M \text{ for some } M \in \mathbb{M}\,\}$, and define $c^{\diamond}\colon \mathbb{M}^{\diamond} \to \mathbb{R}$ by $c^{\diamond}(\pi) = c(M)$ for each $(\pi, M) \in \mathbb{M}^{\diamond} \times \mathbb{M}$ with $\pi^{\wedge} = H_M$.

Lemma 14. *Let $(\mathbb{M}, c)$ be a cost structure, and let W be the real-valued function on Ξ of the form $W(m^{\vee}) = \max_{M\in\mathbb{M}}(\int H_M \,\mathrm{d}m - c(M))$.*

(i) *W is a normalized convex niveloid that has a variational representation $(\mathbb{M}^{\diamond}, c^{\diamond})$.*

(ii) *If $(\mathbb{M}, c)$ is canonical, then $c^{\diamond} = W^*|_{\mathbb{M}^{\diamond}}$.*

(iii) *If $(\mathbb{M}, c)$ is canonical, then $[\arg\max_{M\in\mathbb{M}}(\int H_M \,\mathrm{d}m - c(M))]^{\diamond} = \partial W(m^{\vee}) \cap \mathbb{M}^{\diamond}$ for each $m \in \Delta_{\mathrm{s}}(\Phi)$.*

PROOF. (i) By construction, $W(m^{\vee}) = \max_{\pi\in\mathbb{M}^{\diamond}}(\langle m^{\vee}, \pi\rangle - c^{\diamond}(\pi))$ for each $m \in \Delta_{\mathrm{s}}(\Phi)$. Thus, by Lemma 12 (i) and Lemma 13, $(\mathbb{M}^{\diamond}, c^{\diamond})$ is a variational representation of W. Thus, by Lemma 6 (ii), Lemma 8 (i), and Lemma 10 (iii), W is a convex niveloid. Since c is grounded, so is $c^{\diamond}$. Hence, W is normalized.

(ii) Suppose that $(\mathbb{M}, c)$ is canonical. By Lemmas 12 and 13, $\mathbb{M}^{\diamond}$ and $c^{\diamond}$ are convex, and for each $(\pi, \tilde{\pi}) \in (\mathbb{M}^{\diamond})^2$, if $\langle \xi, \pi\rangle \ge \langle \xi, \tilde{\pi}\rangle$ for each $\xi \in \Xi$, then $c^{\diamond}(\pi) \ge c^{\diamond}(\tilde{\pi})$. Thus, by part (i), $(\mathbb{M}^{\diamond}, c^{\diamond})$ is a canonical variational representation of W. Hence, by Proposition 7, $c^{\diamond} = W^*|_{\mathbb{M}^{\diamond}}$.

(iii) If $(\mathbb{M}, c)$ is canonical, then for each $m \in \Delta_{\mathrm{s}}(\Phi)$, then by part (ii) and Lemma 1 (ii) and (iii), $[\arg\max_{M\in\mathbb{M}}(\int H_M \,\mathrm{d}m - c(M))]^{\diamond} = \arg\max_{\pi\in\mathbb{M}^{\diamond}}(\langle m^{\vee}, \pi\rangle - c^{\diamond}(\pi)) = \arg\max_{\pi\in\mathbb{M}^{\diamond}}(\langle m^{\vee}, \pi\rangle - W^*(\pi)) = \partial W(m^{\vee}) \cap \mathbb{M}^{\diamond}$. □

A binary relation $\succcurlyeq$ on Ξ is

- **Φ-monotone** if $\phi \ge \psi$ implies $\delta[\phi]^{\vee} \succcurlyeq \delta[\psi]^{\vee}$ for each $(\phi, \psi) \in \Phi^2$;
- **attracted to ex post randomization** if $\delta[\lambda\phi + (1-\lambda)\psi]^{\vee} \succcurlyeq [\lambda\delta[\phi] + (1-\lambda)\delta[\psi]]^{\vee}$ for each $(\lambda, (\phi, \psi)) \in [0, 1] \times \Phi^2$.

Lemma 15. *Let W be a normalized convex niveloid on Ξ such that $\succcurlyeq^*_W$ is Φ-monotone and attracted to ex post randomization, and let c be the extended real-valued function on $\mathbb{K}$ of the form $c(M) = \sup_{m \in \Delta_s(\Phi)}(\int H_M \, \mathrm{d}m - W(m^\vee))$. Then, there exists a compact convex subset $\mathbb{M}^*$ of* $\operatorname{dom} c$ *such that*

(i) *for each $(l, m) \in \Delta_s(\Phi)^2$, it follows that $l^\vee \succcurlyeq^*_W m^\vee$ if and only if $\int H_M \, \mathrm{d}l \geq \int H_M \, \mathrm{d}m$ for each $M \in \mathbb{M}^*$;*

(ii) *for each compact convex subset $\mathbb{M}$ of $\mathbb{K}$ with $\mathbb{M}^* \subseteq \mathbb{M} \subseteq \operatorname{dom} c$, the pair $(\mathbb{M}, c|_{\mathbb{M}})$ is a canonical cost structure and $W(m^\vee) = \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}m - c(M))$ for each $m \in \Delta_s(\Phi)$.*

Proof. By construction, $c(M) = W^*(\pi)$ for each $(M, \pi) \in \mathbb{K} \times \mathbb{K}^\diamond$ with $H_M = \pi^\wedge$. By Proposition 6, there exists a weak$*$ compact convex subset Π of $\operatorname{dom} W^*|_\Delta$ such that Π is a multi-expectation representation of $\succcurlyeq^*_W$ and $(\Pi, W^*|_\Pi)$ is a variational representation of W. Since each member of $\Pi^\wedge$ is positively homogeneous and translation equivariant by Lemma 10 (ii) and is monotone and concave by the Φ-monotonicity and attraction to ex post randomization of $\succcurlyeq^*_W$, we have $\Pi \subseteq \Delta_{\mathbb{V}}$. Let $\mathbb{M}^* = \{ M \in \mathbb{K} \mid H_M \in \Pi^\wedge \}$. By Lemma 12 (i) and Lemma 13, $\mathbb{M}^*$ is nonempty, compact, and convex. Since for each $\pi \in \Pi$, there exists $M \in \mathbb{M}$ such that $H_M = \pi^\wedge$, we have (i) and $W(m^\vee) = \max_{M \in \mathbb{M}^*}(\int H_M \, \mathrm{d}m - c(M))$ for each $m \in \Delta_s(\Phi)$. Since $\Pi \subseteq \operatorname{dom} W^*$, we have $\mathbb{M}^* \subseteq \operatorname{dom} c$.

To see (ii), choose any compact convex subset $\mathbb{M}$ of $\mathbb{K}$ with $\mathbb{M}^* \subseteq \mathbb{M} \subseteq \operatorname{dom} c$. Then, for each $m \in \Delta_s(\Phi)$, we have $W(m^\vee) \leq \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}m - c(M))$; since $c(M) \geq \int H_M \, \mathrm{d}m - W(m^\vee)$ for each $M \in \mathbb{M}$, we have $W(m^\vee) \geq \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}m - c(M))$. Since $M \mapsto \int H_M \, \mathrm{d}m - W(m^\vee)$ is mixture linear and continuous for each $m \in \Delta_s(\Phi)$ by Lemma 12 (i), the function $c|_{\mathbb{M}}$ is lower semicontinuous and convex. Since $W^*|_\Pi$ is grounded by Lemma 8 (ii), so is $c|_{\mathbb{M}}$. For each $(M, M') \in \mathbb{M}^2$ with $M \subseteq M'$, since $H_M \geq H_{M'}$, we have $c(M) \geq c(M')$. Thus, $(\mathbb{M}, c)$ is canonical. □

Appendix B. Proofs

B.1. Auxiliary lemmas.

With an abuse of notation, identify $\Delta_s(X)$ with $\{ \delta[p] \mid p \in \Delta_s(X) \}$.

Lemma 16. *If $\succsim$ satisfies regularity, indifference to mixture timing of constant acts, and independence of constant acts, then the restriction of $\succsim$ to $\Delta_s(X)$ is represented by a vNM function.*

Proof. Assume the axioms. Let $(p, q, r) \in \Delta_s(X)^3$ be such that $\delta[p] \sim \delta[q]$. Seeking a contradiction, suppose $\delta[\frac{1}{2}p + \frac{1}{2}r] \succ \delta[\frac{1}{2}q + \frac{1}{2}r]$. Since *indifference to mixture timing of constant acts* implies $\frac{1}{2}\delta[p] + \frac{1}{2}\delta[r] \succ \frac{1}{2}\delta[q] + \frac{1}{2}\delta[r]$, it follows from *independence of constant acts* that $\delta[p] = \frac{1}{2}\delta[p] + \frac{1}{2}\delta[p] \succ \frac{1}{2}\delta[q] + \frac{1}{2}\delta[p]$ and $\frac{1}{2}\delta[p] + \frac{1}{2}\delta[q] \succ \frac{1}{2}\delta[q] + \frac{1}{2}\delta[q] = \delta[q]$. Thus, transitivity implies $\delta[p] \succ \delta[q]$, which is a contradiction. Similarly, $\delta[\frac{1}{2}q + \frac{1}{2}r] \succ \delta[\frac{1}{2}p + \frac{1}{2}r]$ leads to a contradiction. Hence, by completeness, $\delta[\frac{1}{2}p + \frac{1}{2}r] \sim \delta[\frac{1}{2}q + \frac{1}{2}r]$. Therefore, by the mixture space theorem (Herstein and Milnor, 1953), the restriction of $\succsim$ to $\Delta_s(X)$ is represented by a vNM function. □

The next is Lemma 59 of Cerreia-Vioglio et al. (2011).

Lemma 17. *A nondegenerate binary relation on $\Delta_s(X)$ represented by a vNM function u satisfies unboundedness if and only if u is surjective.*

Let $\succsim^u$ be a binary relation on $\Delta_s(\Phi)$. Denote by $\sim^u$ and $\succ^u$ the symmetric and asymmetric parts of $\succsim^u$, respectively. Let $\geq_{\mathrm{FSD}}$ be the first order stochastic dominance relation on $\Delta_s(\Phi)$. The relation $\succsim^u$ is

- **$\varXi$-monotone** if $l^\vee \geq m^\vee$ implies $l \succsim^u m$ for each $(l, m) \in \Delta_s(\Phi)^2$;
- **$\geq_{\mathrm{FSD}}$-monotone** if $l \geq_{\mathrm{FSD}} m$ implies $l \succsim^u m$ for each $(l, m) \in \Delta_s(\Phi)^2$;
- **indifferent to mixture timing of constants** if $\kappa\delta[\lambda\phi + (1-\lambda)t\mathbf{1}_\Omega] + (1-\kappa)m \sim^u \kappa[\lambda\delta[\phi] + (1-\lambda)\delta[t\mathbf{1}_\Omega]] + (1-\kappa)m$ for each $((\kappa,\lambda), m, \phi, t) \in [0,1]^2 \times \Delta_s(\Phi) \times \Phi \times \mathbb{R}$.

Lemma 18. *If $\succsim^u$ is transitive, $\geq_{\mathrm{FSD}}$-monotone, and indifferent to mixture timing of constants, then it is $\varXi$-monotone.*

Proof. Suppose that $\succsim^u$ is transitive, $\geq_{\mathrm{FSD}}$-monotone, and indifferent to mixture timing of constants. For each $((\kappa,\lambda), m, (\alpha,\beta), \psi, (s,t)) \in [0,1]^2 \times \Delta_s(\Phi) \times \mathbb{R}^2_+ \times \Psi \times \mathbb{R}^2$ with $\alpha \leq \beta$, by indifference to mixture timing of constants,

$$\begin{aligned}
&\kappa\big[\lambda\delta[\alpha\psi + s\mathbf{1}_\Omega] + (1-\lambda)\delta[\beta\psi + t\mathbf{1}_\Omega]\big] + (1-\kappa)m \\
&\quad = \kappa\left[\lambda\delta\left[\frac{\alpha}{\beta}(\beta\psi + t\mathbf{1}_\Omega) + \left(1-\frac{\alpha}{\beta}\right)\frac{\beta s - \alpha t}{\beta - \alpha}\mathbf{1}_\Omega\right] + (1-\lambda)\delta[\beta\psi + t\mathbf{1}_\Omega]\right] + (1-\kappa)m \\
&\quad \sim^u \kappa\left[\left[1 - \lambda\left(1-\frac{\alpha}{\beta}\right)\right]\delta[\beta\psi + t\mathbf{1}_\Omega] + \lambda\left(1-\frac{\alpha}{\beta}\right)\delta\left[\frac{\beta s - \alpha t}{\beta-\alpha}\mathbf{1}_\Omega\right]\right] + (1-\kappa)m \\
&\quad \sim^u \kappa\delta\big[[\lambda\alpha + (1-\lambda)\beta]\psi + [\lambda s + (1-\lambda)t]\mathbf{1}_\Omega\big] + (1-\kappa)m,
\end{aligned}$$

and by Lemma 10 (i) and (ii),

$$\begin{aligned}
&\big[\kappa\big[\lambda\delta[\alpha\psi + s\mathbf{1}_\Omega] + (1-\lambda)\delta[\beta\psi + t\mathbf{1}_\Omega]\big] + (1-\kappa)m\big]^\vee \\
&\quad = \big[\kappa\delta\big[[\lambda\alpha + (1-\lambda)\beta]\psi + [\lambda s + (1-\lambda)t]\mathbf{1}_\Omega\big] + (1-\kappa)m\big]^\vee.
\end{aligned}$$

Thus, for each $m \in \Delta_s(\Phi)$, there exist a nonempty finite subset Θ of Ψ and a triplet $(\lambda, \alpha, \tau) \in \mathbb{R}^\Theta_+ \times \mathbb{R}^\Theta_+ \times \mathbb{R}^\Theta$ such that $\sum_{\psi\in\Theta}\lambda(\psi) = 1$, $m \sim^u \sum_{\psi\in\Theta}\lambda(\psi)\delta[\alpha(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega]$, and $m^\vee = [\sum_{\psi\in\Theta}\lambda(\psi)\delta[\alpha(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega]]^\vee$; so for $N = |\Theta| + 1$, again by indifference to mixture timing of constants,

$$m \sim^u \sum_{\psi\in\Theta}\lambda(\psi)\delta[\alpha(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega]$$

$$
\begin{aligned}
&\sim^{\mathrm{u}} \sum_{\psi\in\Theta} \lambda(\psi)\left[\frac{1}{N}\delta[N\alpha(\psi)\psi] + \frac{N-1}{N}\delta\Big[\frac{N}{N-1}\tau(\psi)\mathbf{1}_\Omega\Big]\right] \\
&= \frac{1}{N}\sum_{\psi\in\Theta}\lambda(\psi)\delta[N\alpha(\psi)\psi] + \frac{N-1}{N}\sum_{\psi\in\Theta}\lambda(\psi)\delta\Big[\frac{N}{N-1}\tau(\psi)\mathbf{1}_\Omega\Big] \\
&\sim^{\mathrm{u}} \frac{1}{N}\sum_{\psi\in\Theta}\lambda(\psi)\delta[N\alpha(\psi)\psi] + \frac{N-1}{N}\delta\Big[\frac{N}{N-1}\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega\Big] \\
&= \frac{1}{N}\sum_{\psi\in\Theta}\lambda(\psi)\delta[N\alpha(\psi)\psi] + \frac{N-1}{N}\delta\Big[\frac{N}{N-1}\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega + \frac{N-2}{N-1}0\Big] \\
&\sim^{\mathrm{u}} \frac{1}{N}\sum_{\psi\in\Theta}\lambda(\psi)\delta[N\alpha(\psi)\psi] + \frac{1}{N}\delta\Big[N\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega\Big] + \frac{N-2}{N}\delta[0] \\
&= \frac{1}{N}\sum_{\psi\in\Theta}\big[\lambda(\psi)\delta[N\alpha(\psi)\psi] + (1-\lambda(\psi))\delta[0]\big] + \frac{1}{N}\delta\Big[N\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega\Big] \\
&\sim^{\mathrm{u}} \frac{1}{N}\left[\sum_{\psi\in\Theta}\delta[N\lambda(\psi)\alpha(\psi)\psi] + \delta\Big[N\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega\Big]\right],
\end{aligned}
$$

in which case, by Lemma 10 (ii),

$$
\begin{aligned}
m^\vee &= \Big[\sum_{\psi\in\Theta}\lambda(\psi)\delta[\alpha(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega]\Big]^\vee \\
&= \left\{\frac{1}{N}\left[\sum_{\psi\in\Theta}\delta[N\lambda(\psi)\alpha(\psi)\psi] + \delta\Big[N\sum_{\psi\in\Theta}\lambda(\psi)\tau(\psi)\mathbf{1}_\Omega\Big]\right]\right\}^\vee.
\end{aligned}
$$

Now choose any $(m_1, m_2) \in \Delta_{\mathrm{s}}(\Phi)^2$ with $m_1^\vee \geq m_2^\vee$. By the previous argument, there exists a pair $\langle \Theta, ((\beta_i, \gamma_i))_{i\in\{1,2\}}\rangle$ of a nonempty finite subset of Ψ and a family of pairs of a function from Θ to $\mathbb{R}_+$ and a real number such that for each $i \in \{1, 2\}$,

$$
m_i \sim^{\mathrm{u}} \frac{1}{|\Theta|+1}\Big(\sum_{\psi\in\Theta}\delta[\beta_i(\psi)\psi] + \delta[\gamma_i\mathbf{1}_\Omega]\Big), \qquad m_i^\vee = \Big[\frac{1}{|\Theta|+1}\Big(\sum_{\psi\in\Theta}\delta[\beta_i(\psi)\psi] + \delta[\gamma_i\mathbf{1}_\Omega]\Big)\Big]^\vee.
$$

Since for each $v \in \mathbb{U}$,

$$
\sum_{\psi\in\Psi}(\beta_1(\psi) - \beta_2(\psi))v(\psi) + \gamma_1 - \gamma_2 = (|\Theta|+1)(m_1^\vee(v) - m_2^\vee(v)) \geq 0,
$$

we have

$$
\sum_{\psi\in\Psi}|\beta_1(\psi) - \beta_2(\psi)| \leq \gamma_1 - \gamma_2.
$$

Thus, there exists a function $\tau\colon \Theta \to \mathbb{R}$ such that

$$
\beta_1(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega \geq \beta_2(\psi)\psi \quad \forall \psi \in \Theta \qquad \text{and} \qquad \sum_{\psi\in\Theta}\tau(\psi) \leq \gamma_1 - \gamma_2.
$$

Let
$$\hat{m}_1 = \frac{1}{|\Theta|+1}\left[\sum_{\psi\in\Theta} \delta[\beta_1(\psi)\psi + \tau(\psi)\mathbf{1}_\Omega] + \delta\left[\left(\gamma_1 - \sum_{\psi\in\Theta}\tau(\psi)\right)\mathbf{1}_\Omega\right]\right].$$
Then, $\hat{m}_1 \geq_{\mathrm{FSD}} m_2$. Since for each $(n, m, \phi, (s,t)) \in \mathbb{N}\times\Delta_{\mathrm{s}}(\Phi)\times\Phi\times\mathbb{R}^2$,
$$\begin{aligned}&\frac{1}{n+1}\delta[\phi] + \frac{1}{n+1}\delta[(s+t)\mathbf{1}_\Omega] + \frac{n-1}{n+1}m\\ &\quad\sim^{\mathrm{u}} \frac{1}{n+1}\left(\frac{1}{2}\delta[2\phi] + \frac{1}{2}\delta[0]\right) + \frac{1}{n+1}\left(\frac{1}{2}\delta[2s\mathbf{1}_\Omega] + \frac{1}{2}\delta[2t\mathbf{1}_\Omega]\right) + \frac{n-1}{n+1}m\\ &\quad\sim^{\mathrm{u}} \frac{1}{n+1}\delta[\phi + s\mathbf{1}_\Omega] + \frac{1}{n+1}\delta[t\mathbf{1}_\Omega] + \frac{n-1}{n+1}m,\end{aligned}$$
we have $\hat{m}_1 \sim^{\mathrm{u}} m_1$. Hence, $m_1 \succsim^{\mathrm{u}} m_2$ by the transitivity and $\geq_{\mathrm{FSD}}$-monotonicity of $\succsim^{\mathrm{u}}$. □

The relation $\succsim^{\mathrm{u}}$ is

- **regular** if it is nondegenerate, complete, transitive, and mixture continuous, and $s > t$ implies $\delta[s\mathbf{1}_\Omega] \succ^{\mathrm{u}} \delta[t\mathbf{1}_\Omega]$ for each $(s,t)\in\mathbb{R}^2$;
- **ex ante averse to randomization** if $l \succsim^{\mathrm{u}} m$ implies $l \succsim^{\mathrm{u}} \lambda l + (1-\lambda)m$ for each $(\lambda, (l,m)) \in [0,1]\times\Delta_{\mathrm{s}}(\Phi)^2$;
- **independent of constants** if $\lambda l + (1-\lambda)\delta[s\mathbf{1}_\Omega] \succsim^{\mathrm{u}} \lambda m + (1-\lambda)\delta[s\mathbf{1}_\Omega]$ implies $\lambda l + (1-\lambda)\delta[t\mathbf{1}_\Omega] \succsim^{\mathrm{u}} \lambda m + (1-\lambda)\delta[t\mathbf{1}_\Omega]$ for each $(\lambda, (l,m), (s,t)) \in [0,1]\times\Delta_{\mathrm{s}}(\Phi)^2\times\mathbb{R}^2$.

Lemma 19. *If $\succsim^{\mathrm{u}}$ is regular, $\geq_{\mathrm{FSD}}$-monotone, indifferent to mixture timing of constants, ex ante averse to randomization, and independent of constants, then there exists a normalized convex niveloid W on Ξ such that $m \mapsto W(m^\vee)$ represents $\succsim^{\mathrm{u}}$.*

Proof. Suppose that $\succsim^{\mathrm{u}}$ is regular, $\geq_{\mathrm{FSD}}$-monotone, indifferent to mixture timing of constants, ex ante averse to randomization, and independent of constants. By Lemma 18, it is Ξ-monotone. By the regularity and Ξ-monotonicity of $\succsim^{\mathrm{u}}$, there exists a unique real-valued function W on Ξ such that $\delta[W(m^\vee)\mathbf{1}_\Omega] \sim^{\mathrm{u}} m$ for each $m\in\Delta_{\mathrm{s}}(\Phi)$. Again by Ξ-monotonicity, W is monotone. By regularity, W is normalized, and $m \mapsto W(m^\vee)$ represents $\succsim^{\mathrm{u}}$. For each $(\xi, t)\in\Xi\times\mathbb{R}$, since letting $s = W(\xi + t\mathbf{1}_\Omega) - t$ gives $\frac{1}{2}(2s) + \frac{1}{2}(2t) = W(\frac{1}{2}(2\xi) + \frac{1}{2}(2t)\mathbf{1}_\Omega)$, it follows from the independent of constants that $s = \frac{1}{2}(2s) + \frac{1}{2}0 = W(\frac{1}{2}(2\xi) + \frac{1}{2}0) = W(\xi)$. Hence, W is translation equivariant, so it is a niveloid by Lemma 6 (ii). By the ex ante aversion to randomization of $\succsim^{\mathrm{u}}$, the function W is quasiconvex. Therefore, applying the same argument as in the proof of Lemma 9 of Cerreia-Vioglio et al. (2014) shows that W is convex. □

For each vNM function u and each $P \in \Delta_{\mathrm{s}}(\mathscr{F})$, let P_u be the pushforward of P under $f \mapsto u\circ f$.

Lemma 20. *If u is a surjective vNM function, then $P \mapsto P_u$ from $\Delta_{\mathrm{s}}(\mathscr{F})$ to $\Delta_{\mathrm{s}}(\Phi)$ is surjective and mixture linear.*

PROOF. Let u be a surjective vNM function. The surjectivity follows from the surjectivity of $f \mapsto u \circ f$ from $\mathscr{F}$ to Φ. For each $(\lambda, (P, Q), \Theta) \in [0, 1] \times \Delta_s(\mathscr{F})^2 \times 2^\Phi$, letting $F = \{ f \in \mathscr{F} \mid u \circ f \in \Theta \}$ gives $[\lambda P + (1 - \lambda)Q]_u(\Theta) = [\lambda P + (1 - \lambda)Q](F) = \lambda P(F) + (1 - \lambda)Q(F) = \lambda P_u(\Theta) + (1 - \lambda)Q_u(\Theta)$. Thus, $P \mapsto P_u$ is mixture linear. □

For each binary relation $\succcurlyeq'$ on $\Delta_s(\mathscr{F})$, a binary relation is **more indecisive than** $\succcurlyeq'$ if its graph is included in the graph of $\succcurlyeq'$.

Lemma 21. *Let $\succcurlyeq$ and $\succcurlyeq'$ be binary relations on $\Delta_s(\mathscr{F})$ that have compact convex multi-MEU representations $\mathbb{M}$ and $\mathbb{M}'$, respectively. Then, $\succcurlyeq$ is more indecisive than $\succcurlyeq'$ if and only if $\mathbb{M} \supseteq \mathbb{M}'$.*

PROOF. The necessity of the more indecisiveness of $\succcurlyeq$ follows by definition. For the sufficiency, we show the contrapositive. Suppose $\mathbb{M} \not\supseteq \mathbb{M}'$. Let $M' \in \mathbb{M}' \setminus \mathbb{M}$. By Lemma 11 and Lemma 12 (i), the set $\{ H_M|_\Psi \mid M \in \mathbb{M} \}$ is a compact convex subset of $\mathbb{U}$ that does not contain $H_{M'}|_\Psi$. Thus, by the separation theorem (Aliprantis and Border, 2006, Corollary 5.80), there exists $(\Lambda, t) \in ca(\Psi) \times \mathbb{R}$ such that $\inf_{M \in \mathbb{M}} \int H_M|_\Psi \, d\Lambda > t > \int H_{M'}|_\Psi \, d\Lambda$. Let $a = \max\{\Lambda^+(\Psi), \Lambda^-(\Psi)\}$, which is positive. Define the Borel probability measures l and m on Φ by

$$l(\Theta) = \frac{\Lambda^+(\Theta \cap \Psi)}{2a} + \left(1 - \frac{\Lambda^+(\Psi)}{2a}\right)\mathbf{1}_\Theta(0),$$
$$m(\Theta) = \frac{\Lambda^-(\Theta \cap \Psi)}{2a} + \left(1 - \frac{\Lambda^-(\Psi)}{2a}\right)\mathbf{1}_\Theta\left(\frac{t}{2a - \Lambda^-(\Psi)}\mathbf{1}_\Omega\right),$$

where $\mathbf{1}_\Theta$ is the indicator function of Θ on Φ for each Borel subset Θ of Φ. Then, for each $\tilde{M} \in \mathbb{M} \cup \{M'\}$, by the normalizedness of $H_{\tilde{M}}$,

$$\int H_{\tilde{M}} \, dl = \int H_{\tilde{M}} \mathbf{1}_\Psi \, dl + \int H_{\tilde{M}} \mathbf{1}_{\Psi^c} \, dl = \frac{1}{2a} \int H_{\tilde{M}} \, d\Lambda^+,$$
$$\begin{aligned}\int H_{\tilde{M}} \, dm &= \int H_{\tilde{M}} \mathbf{1}_\Psi \, dm + \int H_{\tilde{M}} \mathbf{1}_{\Psi^c} \, dm \\ &= \frac{1}{2a} \int H_{\tilde{M}} \, d\Lambda^- + \left(1 - \frac{\Lambda^+(\Psi)}{2a}\right) m\left(\left\{\frac{t}{2a - \Lambda^-(\Psi)}\mathbf{1}_\Omega\right\}\right) = \frac{1}{2a}\left(\int H_{\tilde{M}} \, d\Lambda^- + t\right).\end{aligned}$$

Thus, $\int H_M \, dl > \int H_M \, dm$ for each $M \in \mathbb{M}$, and $\int H_{M'} \, dl < \int H_{M'} \, dm$. Since $\Delta_s(\Phi)$ is dense in the space of Borel probability measures on Φ (Aliprantis and Border, 2006, Theorem 15.10), it is without loss of generality to assume that l and m are finitely supported. Thus, $l \succcurlyeq m$ and $l \not\succcurlyeq' m$, so $\succcurlyeq$ is not more indecisive than $\succcurlyeq'$. □

Lemma 22. *Let $\mathbb{M}$ and $\mathbb{M}'$ be compact convex subsets of $\mathbb{K}$. Then, $\mathbb{M} \trianglerighteq \mathbb{M}'$ if and only if $\max_{M \in \mathbb{M}} \int H_M \, dm \geq \max_{M' \in \mathbb{M}'} \int H_{M'} \, dm$ for each $m \in \Delta_s(\Phi)$.*

PROOF. If $\mathbb{M} \trianglerighteq \mathbb{M}'$, then for each $m \in \Delta_s(\Phi)$, letting $\bar{M}' \in \arg\max_{M' \in \mathbb{M}'} \int H_M \, dm$ gives $\bar{M} \subseteq \bar{M}'$ for some $\bar{M} \in \mathbb{M}$, so $\max_{M' \in \mathbb{M}'} \int H_{M'} \, dm = \int H_{\bar{M}'} \, dm \leq \int H_{\bar{M}} \, dm \leq \max_{M \in \mathbb{M}} \int H_M \, dm$ by Lemma 12 (ii). For the converse, we show the contrapositive. Suppose that there exists

$M' \in \mathbb{M}'$ such that $M \not\subseteq M'$ for each $M \in \mathbb{M}$. By Lemmas 11 and 12, the set $\{ H_M|_\Psi - H_{M'}|_\Psi \mid M \in \mathbb{M} \}$ is a compact convex subset of $\mathbb{U}$ that is disjoint from $\mathrm{C_b}(\Psi)_+$. Thus, by the separation theorem (Aliprantis and Border, 2006, Theorem 5.79), there exists $(\Lambda, t) \in ca(\Psi) \times \mathbb{R}$ such that $\inf_{h \in \mathrm{C_b}(\Psi)_+} \int h \, \mathrm{d}\Lambda > t > \max_{M \in \mathbb{M}} \int H_M|_\Psi \, \mathrm{d}\Lambda - \int H_{M'}|_\Psi \, \mathrm{d}\Lambda$, which implies $(\Lambda, t) \in ca(\Psi)_+ \times (-\mathbb{R}_{++})$. Since $\Lambda \neq 0$, we have $\Lambda(\Psi) > 0$. Define the Borel probability measure m on Φ by $m(\Theta) = \Lambda(\Theta \cap \Psi)/\Lambda(\Psi)$. Then, $\int H_{M'} \, \mathrm{d}m > \max_{M \in \mathbb{M}} \int H_M \, \mathrm{d}m$. Since $\Delta_\mathrm{s}(\Phi)$ is dense in the space of Borel probability measures on Φ (Aliprantis and Border, 2006, Theorem 15.10), it is without loss of generality to assume that m is finitely supported. □

B.2. **Proof of Theorem 1.** For the necessity of the axioms, suppose that $\succsim$ has a costly ambiguity perception representation $\langle u, (\mathbb{M}, c) \rangle$. The necessity of *regularity* is routine. By Lemma 17, *unboundedness* holds. Define $W \colon \Xi \to \mathbb{R}$ by $W(P_u^\vee) = \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}P_u - c(M))$, which is a convex niveloid by Lemma 14 (i). Since $P \mapsto P_u^\vee$ is mixture linear by Lemma 10 (i) and Lemma 20, the function $P \mapsto W(P_u^\vee)$ is convex. Thus, *ex ante aversion to randomization* is satisfied. Since $W([\lambda P + (1-\lambda)\delta[p]]_u^\vee) = W(\lambda P_u^\vee + (1-\lambda)u(p)\mathbf{1}_\mathbb{U}) = W(\lambda P_u^\vee) + (1-\lambda)u(p)$ for each $(\lambda, P, p) \in [0,1] \times \Delta_\mathrm{s}(\mathscr{F}) \times \Delta_\mathrm{s}(X)$, *independence of constant acts* holds. The remained axioms, *FSD*, *attraction to ex post randomization*, and *indifference to mixture timing of constant acts*, follow from the monotonicity, concavity, and translation equivariance of the support functions, respectively.

For the sufficiency, assume all the axioms. By Lemmas 16 and 17, the restriction of $\succsim$ to $\Delta_\mathrm{s}(X)$ is represented by a surjective vNM function u. Define the relation $\succsim^\mathrm{u}$ on $\Delta_\mathrm{s}(\Phi)$ by $P_u \succsim^\mathrm{u} Q_u$ if $P \succsim Q$. By Lemma 20, $\succsim^\mathrm{u}$ is regular, $\geq_\mathrm{FSD}$-monotone, indifferent to mixture timing of constants, ex ante averse to randomization, and independent of constants. Thus, by Lemma 19, there exists a normalized convex niveloid W on Ξ such that $P \mapsto W(P_u^\vee)$ represents $\succsim$. By *FSD* and *attraction to ex post randomization*, $\succcurlyeq_W^*$ is Φ-monotone and attracted to ex post randomization. Hence, by Lemma 15, there exists a cost structure $(\mathbb{M}, c)$ such that for each $P \in \Delta_\mathrm{s}(\mathscr{F})$,

$$W(P_u^\vee) = \max_{M \in \mathbb{M}} \left(\int H_M \, \mathrm{d}P_u - c(M) \right) = \max_{M \in \mathbb{M}} \left[\int \left(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu \right) \mathrm{d}P(f) - c(M) \right],$$

which shows that $\langle u, (\mathbb{M}, c) \rangle$ is a costly ambiguity perception representation of $\succsim$. □

B.3. **Proof of Proposition 1.** By Lemma 21, $\succsim^*$ has at most one compact convex multi-MEU representation. Let $\langle u, (\mathbb{M}, c) \rangle$ be a costly ambiguity perception representation of $\succsim$. Define $W \colon \Xi \to \mathbb{R}$ by $W(m^\vee) = \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}m - c(M))$. Then, for each $(P, Q) \in \Delta_\mathrm{s}(\mathscr{F})^2$, it follows that $P_u^\vee \succcurlyeq_W^* Q_u^\vee$ if and only if $P \succsim^* Q$. By Lemma 14 (i), W is a normalized convex niveloid. By the monotonicity and concavity of support functions, $\succcurlyeq_W^*$ is Φ-monotone and attracted to ex post randomization. Thus, by Lemma 15, $\succsim^*$ has a compact convex multi-MEU representation. □

B.4. **Proof of Theorem 2.** (i) Suppose that $\langle u, (\mathbb{M}, c)\rangle$ is a canonical costly ambiguity perception representation of $\succsim$. Define $W\colon \Xi \to \mathbb{R}$ by $W(m^\vee) = \max_{M\in\mathbb{M}}(\int H_M \, \mathrm{d}m - c(M))$. Then, $W(P_u^\vee) = u(\bar{P})$ for each $P \in \Delta_s(\mathscr{F})$. Define the relation $\succsim'$ on $\Delta_s(\mathscr{F})$ by $P \succsim' Q$ if

$$\int \Big(\min_{\mu\in M} \int u \circ f \, \mathrm{d}\mu\Big) \, \mathrm{d}P(f) \geq \int \Big(\min_{\mu\in M} \int u \circ f \, \mathrm{d}\mu\Big) \, \mathrm{d}Q(f) \quad \forall M \in \mathbb{M}.$$

For each $(P, Q) \in \Delta_s(\mathscr{F})^2$, if $P \succsim' Q$, then for each $(\lambda, R) \in (0,1] \times \Delta_s(\mathscr{F})$, every $\bar{M} \in \arg\max_{M\in\mathbb{M}}\{\int H_M \, \mathrm{d}[\lambda Q + (1-\lambda)R]_u - c(M)\}$ satisfies

$$\begin{aligned} W([\lambda P + (1-\lambda)R]_u^\vee) &\geq \int H_{\bar{M}} \, \mathrm{d}[\lambda P + (1-\lambda)R]_u - c(\bar{M}) \\ &\geq \int H_{\bar{M}} \, \mathrm{d}[\lambda Q + (1-\lambda)R]_u - c(\bar{M}) = W([\lambda Q + (1-\lambda)R]_u^\vee), \end{aligned}$$

so $P \succsim^* Q$. That is, $\succsim'$ is more indecisive than $\succsim^*$. Thus, by Lemma 21, $\mathbb{M} \supseteq \mathbb{M}^*$. Since $(\mathbb{M}, c)$ is canonical, we have $c^\diamond = W^*|_{\mathbb{M}^\diamond}$ by Lemma 14 (ii). Hence, $c(M) = \sup_{\xi\in\Xi}(\langle \xi, \pi\rangle - W(\xi)) = \sup_{m\in\Delta_s(\Phi)}(\int H_M \, \mathrm{d}m - W(m^\vee)) = c^\star_{\succsim,u}(M)$ for each $(M, \pi) \in \mathbb{M} \times \mathbb{M}^\diamond$ with $H_M = \pi^\wedge$. Since c is real-valued, we have $\mathbb{M} \subseteq \operatorname{dom} c^\star_{\succsim,u}$.

For the converse, suppose that $\mathbb{M}^* \subseteq \mathbb{M} \subseteq \operatorname{dom} c^\star_{\succsim,u}$ and $c = c^\star_{\succsim,u}|_{\mathbb{M}}$. Define $W\colon \Xi \to \mathbb{R}$ by $W(P_u^\vee) = u(\bar{P})$. Then, for each $(P, Q) \in \Delta_s(\mathscr{F})^2$, it follows that $P_u^\vee \succcurlyeq_W^* Q_u^\vee$ if and only if $P \succsim^* Q$. Since there exists a cost structure $(\bar{\mathbb{M}}, \bar{c})$ such that $u(\bar{P}) = \max_{M\in\bar{\mathbb{M}}}(\int H_M \, \mathrm{d}P_u - \bar{c}(M))$ for each $P \in \Delta_s(\mathscr{F})$, it follows from Lemma 14 (i) that W is a convex niveloid. By the monotonicity and concavity of support functions, $\succcurlyeq_W^*$ is Φ-monotone and attracted to ex post randomization. Thus, by Proposition 1 and Lemma 15, $(\mathbb{M}, c)$ is a canonical cost structure such that $u(\bar{P}) = W(P_u^\vee) = \max_{M\in\mathbb{M}}(\int H_M \, \mathrm{d}P_u^\vee - c(M))$ for each $P \in \Delta_s(\mathscr{F})$.

(ii) Suppose that $\langle u, (\mathbb{M}, c)\rangle$ is a canonical costly ambiguity perception representation of $\succsim$ such that $\mathbb{M}$ is $\supseteq$-increasing. By part (i), $c = c^\star_{\succsim,u}|_{\mathbb{M}}$. Define $\hat{c}\colon \mathbb{K} \to [0, \infty]$ by $\hat{c}(M) = c(M)$ if $M \in \mathbb{M}$ and $\hat{c}(M) = \infty$, and define $W\colon \Xi \to \mathbb{R}$ by $W(m^\vee) = \max_{M\in\mathbb{K}}(\int H_M \, \mathrm{d}m - \hat{c}(M))$. Then, by Lemmas 12 and 13 and the $\supseteq$-increasingness of $\mathbb{M}$, the pair $(\mathbb{K}^\diamond, \hat{c}^\diamond)$ is a canonical variational representation of W. Thus, by Proposition 7, $\hat{c}(M) = \hat{c}^\diamond(\pi) = W^*|_{\mathbb{K}^\diamond}(\pi) = c^\star_{\succsim,u}(M)$ for each $(M, \pi) \in \mathbb{K} \times \mathbb{K}^\diamond$ with $H_M = \pi^\wedge$. Hence, $\mathbb{M} = \operatorname{dom} \hat{c} = \operatorname{dom} c^\star_{\succsim,u}$.

For the converse, suppose $\mathbb{M} = \operatorname{dom} c^\star_{\succsim,u}$ and $c = c^\star_{\succsim,u}|_{\mathbb{M}}$. By part (i), $\langle u, (\mathbb{M}, c)\rangle$ is a canonical costly ambiguity perception representation of $\succsim$. For each $(M, M') \in \mathbb{K}^2$, if $M \in \mathbb{M}$ and $M \subseteq M'$, then $c(M) \geq c(M')$, so $M' \in \mathbb{M}$. Thus, $(\mathbb{M}, c)$ is $\supseteq$-increasing. □

B.5. **Proof of Proposition 2.** Let $i \in \{1, 2\}$. Define $W_i\colon \Xi \to \mathbb{R}$ by $W_i(m^\vee) = \max_{M\in\mathbb{M}_i}(\int H_M \, \mathrm{d}m - c_i(M))$. For each $(P, Q) \in \Delta_s(\mathscr{F})^2$, it follows from Lemma 14 (i) and (iii) that $\mathscr{C}_i(P) \cap \mathscr{C}_i(Q) \neq \emptyset$ if and only if $\partial W_i(P_{u_i}^\vee) \cap \partial W_i(Q_{u_i}^\vee) \cap \mathbb{M}_i^\diamond \neq \emptyset$, which is equivalent to $W_i([\lambda P + (1-\lambda)Q]_{u_i}^\vee) = W_i(\lambda P_{u_i}^\vee + (1-\lambda)Q_{u_i}^\vee) = \lambda W_i(P_{u_i}^\vee) + (1-\lambda)W_i(Q_{u_i}^\vee) = \lambda u_i(\bar{P}_i) + (1-\lambda)u_i(\bar{Q}_i)$ for each $\lambda \in [0, 1]$ by Lemma 9 (i). Thus, the desired equivalence is obtained. □

B.6. **Proof of Proposition 3.** DM 1 is more tolerant of ambiguity than DM 2 if and only if $u_1 \approx u_2$ and $u_1(\bar{P}_1) \geq u_1(\bar{P}_2)$, while the inequality is equivalent to $c^{\star}_{\succsim_1, u_1} \leq c^{\star}_{\succsim_2, u_1}$. □

B.7. **Proof of Proposition 4.** If DM 1 has higher filtering incentives than DM 2, then $u_1 \approx u_2$. Thus, assume without loss of generality $u_1 = u_2$. Let $u = u_1$. For each $i \in \{1, 2\}$, define $W_i \colon \Xi \to \mathbb{R}$ by $W_i(m^\vee) = \max_{M \in \mathbb{M}_i}(\int H_M \, \mathrm{d}m - c_i(M))$, and define $U_i \colon \Delta_{\mathrm{s}}(\mathscr{F}) \to \mathbb{R}$ by $U_i(P) = W_i(P_u^\vee)$.

Suppose that $\mathscr{C}_1(P) \trianglerighteq \mathscr{C}_2(P)$ for each $P \in \Delta_{\mathrm{s}}(\mathscr{F})$. By Lemma 22, for each $(P, Q) \in \Delta_{\mathrm{s}}(\mathscr{F})^2$,

$$\max_{M \in \mathscr{C}_1(P)} \int \Big(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu\Big) \, \mathrm{d}Q(f) \geq \max_{M \in \mathscr{C}_2(P)} \int \Big(\min_{\mu \in M} \int u \circ f \, \mathrm{d}\mu\Big) \, \mathrm{d}Q(f).$$

Choose any $(\lambda, (P, Q), p) \in [0, 1] \times \Delta_{\mathrm{s}}(\mathscr{F})^2 \times \Delta_{\mathrm{s}}(X)$. Define $R \colon [0, 1] \to \Delta_{\mathrm{s}}(\mathscr{F})$ by $R(t) = \lambda[tP + (1 - t)\delta[p]] + (1 - \lambda)Q$. Let $\tilde{P} \in \Delta_{\mathrm{s}}(\mathscr{F})$ be such that $\int H_M \, \mathrm{d}\tilde{P}_u = \int H_M \, \mathrm{d}P_u - u(p)$ for each $M \in \mathbb{M}$. For each $t \in [0, 1]$, let $(\bar{M}_1(t), \bar{M}_2(t)) \in \mathscr{C}_1(R(t)) \times \mathscr{C}_2(R(t))$ be such that $\int H_{\bar{M}_1(t)} \, \mathrm{d}\tilde{P}_u \geq \int H_{\bar{M}_2(t)} \, \mathrm{d}\tilde{P}_u$. Then, since for each $i \in \{1, 2\}$, the envelope theorem (Milgrom and Segal, 2002, Theorem 2) implies

$$U_i(\lambda P + (1 - \lambda)Q) - U_i(\lambda\delta[p] + (1 - \lambda)Q) = U_i(R(1)) - U_i(R_i(0)) = \lambda \int_0^1 \Big(\int H_{\bar{M}_i(t)} \, \mathrm{d}\tilde{P}_u\Big) \, \mathrm{d}t,$$

we have $U_1(\lambda P + (1 - \lambda)Q) - U_1(\lambda\delta[p] + (1 - \lambda)Q) \geq U_2(\lambda P + (1 - \lambda)Q) - U_2(\lambda\delta[p] + (1 - \lambda)Q)$. Thus, DM 1 has higher filtering incentives than DM 2.

For the converse, suppose that DM 1 has higher filtering incentives than DM 2. Choose any $(P, m) \in \Delta_{\mathrm{s}}(\mathscr{F}) \times \Delta_{\mathrm{s}}(\Phi)$. Since for each $i \in \{1, 2\}$, by Lemma 14 (i) and (iii) and Lemma 9 (ii),

$$\max_{M \in \mathscr{C}_i(P)} \int H_M \, \mathrm{d}m = \max_{\pi \in \partial W_i(P_u^\vee) \cap \mathbb{M}_i^\circ} \langle m^\vee, \pi \rangle = \mathrm{D}^+_{m^\vee} W_i(P_u^\vee),$$

it suffices to show $\mathrm{D}^+_{m^\vee} W_1(P_u^\vee) \geq \mathrm{D}^+_{m^\vee} W_2(P_u^\vee)$ by Lemma 22. Let $\tilde{P} \in \Delta_{\mathrm{s}}(\mathscr{F})$ be such that $\tilde{P}_u^\vee = 2P_u^\vee$. For each $\lambda \in (0, 1]$, let $\tilde{Q}(\lambda) \in \Delta_{\mathrm{s}}(\mathscr{F})$ be such that $\tilde{Q}(\lambda)_u^\vee = 2\lambda m^\vee$, and let $p(\lambda) \in \Delta_{\mathrm{s}}(X)$ be such that $u(p(\lambda)) = 2(W_2(P_u^\vee + \lambda m^\vee) - W_2(P_u^\vee))$; then, since

$$U_2\Big(\frac{1}{2}\tilde{P} + \frac{1}{2}\tilde{Q}(\lambda)\Big) - U_2\Big(\frac{1}{2}\tilde{P} + \frac{1}{2}\delta[p(\lambda)]\Big) = W_2(P_u^\vee + \lambda m^\vee) - W_2(P_u^\vee) - \frac{1}{2}u(p(\lambda)) = 0,$$

we have

$$W_1(P_u^\vee + \lambda m^\vee) - W_1(P_u^\vee) - \frac{1}{2}u(p(\lambda)) = U_1\Big(\frac{1}{2}\tilde{P} + \frac{1}{2}\tilde{Q}(\lambda)\Big) - U_1\Big(\frac{1}{2}\tilde{P} + \frac{1}{2}\delta[p(\lambda)]\Big) \geq 0.$$

Thus, $\mathrm{D}^+_{m^\vee} W_1(P_u^\vee) \geq \mathrm{D}^+_{m^\vee} W_2(P_u^\vee)$. □

B.8. **Proof of Corollary 2.** In each statement, the necessity of the axiom is routine. We show only the sufficiency. Let $\langle u, (\mathbb{M}, c) \rangle$ be a costly ambiguity perception representation of $\succsim$. By Proposition 1 and Theorem 2 (i), we may assume that $\mathbb{M}$ is a multi-MEU representation of $\succsim^*$. Define $U \colon \Delta_{\mathrm{s}}(\mathscr{F}) \to \mathbb{R}$ by $U(P) = \max_{M \in \mathbb{M}}(\int H_M \, \mathrm{d}P_u - c(M))$. Let $\sim^*$ be the symmetric part of $\succsim^*$. Observe that for each $(\lambda, (f, g)) \in [0, 1] \times \mathscr{F}^2$, if $\delta[\lambda f + (1 - \lambda)g] \sim^* \lambda\delta[f] + (1 - \lambda)\delta[g]$, then $\lambda H_M(u \circ f) + (1 - \lambda)H_M(u \circ g) = H_M(\lambda(u \circ f) + (1 - \lambda)(u \circ g))$ for each $M \in \mathbb{M}$.

(i) Assume *indifference to mixture timing of comonotonic acts*. By the above observation, for each $(\phi, \psi) \in \Phi^2$ if $(\phi(\omega) - \phi(\omega'))(\psi(\omega) - \psi(\omega')) \geq 0$ for each $(\omega, \omega') \in \Omega^2$, then $H_M(\phi) + H_M(\psi) = H_M(\phi + \psi)$ for each $M \in \mathbb{M}$. Thus, by the Theorem and Proposition 3 of Schmeidler (1986), every member of $\mathbb{M}$ is the core of a convex capacity.

(ii) Assume *indifference to mixture timing*. For each $M \in \mathbb{M}$, the function H_M is mixture linear by the above observation, so it is linear. Thus, by Theorem 5.54 of Aliprantis and Border (2006), every member of $\mathbb{M}$ is a singleton.

(iii) Assume *strong independence of constant acts*. Define $\mathscr{C}\colon \Delta_s(\mathscr{F}) \rightrightarrows \mathbb{M}$ by $\mathscr{C}(P) = \arg\max_{M \in \mathbb{M}}(\int H_M \, dP_u - c(M))$. We show that $\mathscr{C}(P) \cap c^{-1}(\{0\}) \neq \emptyset$ for each $P \in \Delta_s(\mathscr{F})$, in which case $(u, \mathbb{M} \cap c^{-1}(\{0\}))$ is a dual-self expected utility representation of $\succsim$. Choose any $P \in \Delta_s(\mathscr{F})$. Let $p \in \Delta_s(X)$ be such that $\delta[p] \sim P$, let $Q = \frac{1}{2}P + \frac{1}{2}\delta[p]$, and let $M \in \mathscr{C}(Q)$. Then, $c(M) \geq \int H_M \, dP_u - U(P)$. Since $P \sim Q$ by *strong independence of constant acts*, we have

$$U(P) = U(Q) = \frac{1}{2}\int H_M \, dP_u + \frac{1}{2}u(p) - c(M) = \frac{1}{2}\int H_M \, dP_u + \frac{1}{2}U(P) - c(M),$$

so $c(M) = (\int H_M \, dP_u - U(P))/2$. Thus, $\int H_M \, dP_u - U(P) = 0$, which implies $M \in \mathscr{C}(P) \cap c^{-1}(\{0\})$. □

B.9. **Proof of Proposition 5.** Assume *regularity*, *attraction to ex post randomization*, *indifference to mixture timing of constant acts*, *ex ante attraction to randomization*, and *strong independence of constant acts*. Choose any $(\lambda, (f, g)) \in (0,1) \times \mathscr{F}^2$. If $\delta[f] \succsim \delta[g]$, then $\delta[\lambda f + (1-\lambda)g] \succsim \lambda\delta[f] + (1-\lambda)\delta[g] \succsim \delta[g]$ by *attraction to ex post randomization* and *ex ante attraction to randomization*. Also, by *indifference to mixture timing of constant acts* and *strong independence of constant acts*, $\delta[f] \succsim \delta[g]$ if and only if $\delta[\lambda f + (1-\lambda)p] \sim \lambda\delta[f] + (1-\lambda)\delta[p] \succsim \lambda\delta[g] + (1-\lambda)\delta[p] \sim \delta[\lambda g + (1-\lambda)p]$ for each $p \in \Delta_s(X)$. □

Appendix C. Machina's examples

We explore the implications of the dual-self expected utility model on the behavior in Machina's examples. As in Section 5, for each example, identify each (i, j) with the probability distribution over colors such that the probability of drawing blue is i and drawing green is j.

The dual-self expected utility model can explain the pattern $f_6 \succ f_5$ and $f_7 \succ f_8$ in the reflection example. For example, let $M_1 = \{1/4\} \times [0, 1/2]$, let $M_2 = [0, 1/2] \times \{1/4\}$, and let $\mathbb{M} = \{M_1, M_2\}$. Let U be the utility function over acts corresponding to the set $\mathbb{M}$ of possible ambiguity perceptions. Then,

$$U(f_5) = \max_{k \in \{1,2\}} \min_{(b,g) \in M_k} 100\left(\frac{1}{2} + b + g\right) = 75, \qquad U(f_6) = \max_{k \in \{1,2\}} \min_{(b,g) \in M_k} 100\left(\frac{1}{2} + 2g\right) = 100,$$

$$U(f_7) = \max_{k \in \{1,2\}} \min_{(b,g) \in M_k} 100\left(2b + \frac{1}{2}\right) = 100, \qquad U(f_8) = \max_{k \in \{1,2\}} \min_{(b,g) \in M_k} 100\left(b + g + \frac{1}{2}\right) = 75,$$

so $f_6 \succ f_5$ and $f_7 \succ f_8$. In this example, M_1 corresponds to the filtering of ambiguity perception only on blue balls, and M_2 to green balls. Since the DM can exert filtering only either on blue or green, the "more ambiguous" acts f_5 and f_8 is evaluated lower. At the same time, this parameterization can explain the typical Ellsberg-pattern. The utility f_9 and f_{10} of each act is given by

$$U(f_9) = 100 \times \frac{1}{2} = 50, \qquad U(f_{10}) = \max_{k\in\{1,2\}} \min_{(b,g)\in M_k} 100(b+g) = 25,$$

which shows $f_9 \succ f_{10}$.

In contrast, it fails to accommodate $f_1 \succ f_2$ and $f_4 \succ f_3$ in the 50–51 example. Suppose that $\succsim$ has a dual-self expected utility representation $(u, \mathbb{M})$. Define auxiliary acts g, h, and p as described in Table 4. Then,

$$f_1 = \frac{1}{3}g + \frac{2}{3}p, \qquad f_2 = \frac{1}{3}h + \frac{2}{3}p, \qquad f_3 = \frac{2}{3}g + \frac{1}{3}h, \qquad f_4 = \frac{1}{3}g + \frac{2}{3}h.$$

Suppose $f_1 \succ f_2$. Since $\succsim$ satisfies the *certainty independence* axiom (Chandrasekher et al., 2022, Theorem 1), we have $g \succ h$. Thus, since for each $\lambda \in [0,1]$,

$$\begin{aligned} U(\lambda g + (1-\lambda)h) &= \max_{M\in\mathbb{M}} \min_{\mu\in M} \left[\lambda \int u\circ g \,\mathrm{d}\mu + (1-\lambda)\int u\circ h\,\mathrm{d}\mu\right] \\ &= \max_{M\in\mathbb{M}} \min_{\mu\in M} \left[\lambda\left(300\times\frac{50}{101}\right) + (1-\lambda)\int u\circ h\,\mathrm{d}\mu\right] \\ &= \lambda\left(300\times\frac{50}{101}\right) + (1-\lambda)\max_{M\in\mathbb{M}} \min_{\mu\in M}\int u\circ h\,\mathrm{d}\mu = \lambda U(g) + (1-\lambda)U(h), \end{aligned}$$

we have $f_3 \succ f_4$.

	50 balls		51 balls	
	Red	Blue	Green	Purple
g	300	300	0	0
h	300	0	300	0
p	150	150	150	150

TABLE 4. Auxiliary acts.

REFERENCES

Aliprantis, C. D. and K. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*: Springer, 3rd ed. [21, 22, 23, 24, 25, 30, 31, 34]

Anscombe, F. J. and R. J. Aumann (1963). "A definition of subjective probability." *Annals of Mathematical Statistics*, 34, 199–205. [2]

Baillon, A., O. L'Haridon, and L. Placido (2011). "Ambiguity models and the Machina paradoxes." *American Economic Review*, 101, 1547–1560. [3, 12, 13, 18]

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011). "Uncertainty averse preferences." *Journal of Economic Theory*, 146, 1275–1330. [3, 27]

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2014). "Niveloids and their extensions: Risk measures on small domains." *Journal of Mathematical Analysis and Applications*, 413, 343–360. [8, 20, 29]

——— (2015). "The structure of variational preferences." *Journal of Mathematical Economics*, 57, 12–19. [8]

Chandrasekher, M., M. Frick, R. Iijima, and Y. Le Yaouanq (2022). "Dual-self representations of ambiguity preferences." *Econometrica*, 90, 1029–1061. [3, 4, 16, 35]

de Oliveira, H., T. Denti, M. Mihm, and K. Ozbek (2017). "Rationally inattentive preferences and hidden information costs." *Theoretical Economics*, 12, 621–654. [4, 7, 10, 22]

Dillenberger, D. and U. Segal (2015). "Recursive ambiguity and Machina's examples." *International Economic Review*, 56, 55–61. [4]

Ellsberg, D. (1961). "Risk, ambiguity, and the Savage axioms." *Quarterly Journal of Economics*, 19, 643–669. [1]

Ergin, H. and T. Sarver (2010). "A unique costly contemplation representation." *Econometrica*, 78, 1285–1339. [4, 7]

——— (2015). "Hidden actions and preferences for timing of resolution of uncertainty." *Theoretical Economics*, 10, 489–541. [4]

Gale, D. (1967). "A geometric duality theorem with economic applications." *Review of Economic Studies*, 34, 19–24. [19]

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004). "Differentiating ambiguity and ambiguity attitude." *Journal of Economic Theory*, 118, 133–173. [3]

Ghirardato, P. and M. Marinacci (2002). "Ambiguity made precise: A comparative foundation." *Journal of Economic Theory*, 102, 251–289. [11]

Gilboa, I. and D. Schmeidler (1989). "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18, 141–153. [2]

Gul, F. and W. Pesendorfer (2014). "Expected uncertain utility theory." *Econometrica*, 82, 1–39. [5]

He, Y. (2021). "Revisiting Ellsberg's and Machina's paradoxes: A two-stage evaluation model under ambiguity." *Management Science*, 67, 6897–6914. [5]

Herstein, I. N. and J. Milnor (1953). "An axiomatic approach to measurable utility." *Econometrica*, 21, 291–297. [27]

Ke, S. and Q. Zhang (2020). "Randomization and ambiguity aversion." *Econometrica*, 88, 1159–1195. [2, 4, 9, 16, 17]

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). "A smooth model of decision making under ambiguity." *Econometrica*, 73, 1849–1892. [3]

L'Haridon, O. and L. Placido (2010). "Betting on Machina's reflection example: An experiment on ambiguity." *Theory and Decision*, 69, 375–393. [13]

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498. [3, 18]

Machina, M. J. (2009). "Risk, ambiguity, and the rank-dependence axioms." *American Economic Review*, 99, 385–392. [1, 3, 4, 12, 13]

Milgrom, P. and I. Segal (2002). "Envelope theorems for arbitrary choice sets." *Econometrica*, 70, 583–601. [33]

Payró, F., N. Takeoka, and J. Xia (2025). "Optimal ambiguity perception." Unpublished manuscript. [3, 4, 15, 16]

Pennesi, D. (2015). "Costly information acquisition and the temporal resolution of uncertainty." *Journal of Mathematical Economics*, 60, 115–122. [4, 19]

Schmeidler, D. (1986). "Integral representation without additivity." *Proceedings of the American Mathematical Society*, 97, 255–261. [34]

——— (1989). "Subjective probability and expected utility without additivity." *Econometrica*, 57, 571. [3]

Sinander, L. (2024). "Optimism, overconfidence, and moral hazard." *Journal of Political Economy: Microeconomics*, forthcoming. [3, 15, 16]

Siniscalchi, M. (2009). "Vector expected utility and attitudes toward variation." *Econometrica*, 77, 801–855. [4]

Wang, F. (2022). "Rank-dependent utility under multiple priors." *Management Science*, 68, 8166–8183. [4]

Zălinescu, C. (2002). *Convex Analysis in General Vector Spaces*: World Scientific. [18, 19, 24]