New Economic School

Working Paper

No 291

NES Working Paper series

SECOND-ORDER REPRESENTATIONS: A BAYESIAN APPROACH

Ozgur Evren

January 2024

Second-Order Representations: A Bayesian Approach*

Özgür Evren[†]

January 10, 2024

Abstract

For choice problems under ambiguity, I provide a behavioral characterization of a decision maker who holds a second-order belief and updates it in a Bayesian fashion in response to new information concerning the true distribution of the states. The model features a unique second-order belief that can be elicited from choice data and is quite comprehensive in terms of ambiguity attitudes and risk preferences. Special versions, such as the smooth ambiguity model or the recursive non-expected utility model, are easily characterized by additional assumptions on compound-risk preferences. Thereby, the model provides a testing ground to compare and contrast these well-known representations as well as alternative specifications that may be of interest. To illustrate potential benefits of alternative specifications, I provide a detailed analysis of a rank-dependent extension of the smooth ambiguity model.

Keywords. Ambiguity Aversion and Seeking, Ellsberg Paradox, Second-Order Belief, Probabilistic Sophistication, Bayesian Updating, Compound Risk

JEL Classification. D81, D83

*I thank A. Baillon, D. Dillenberger, D. Freeman, Y. Halevy, I. Kopylov, F. Maccheroni, P. Ortoleva, E. Ozdenoren, A. Savochkin, P. Wakker, and the participants of RUD 2020 for useful discussions.

[†]New Economic School, Skolkovo Innovation Center, Nobelya Street, 3, Moscow 121205, Russia. Email: oevren@nes.ru

1. Introduction

Subjective expected utility models consist of two components: An expected utility operator and a probability distribution on the states of nature that represents the decision maker's (DM) subjective prior belief. Machina and Schmeidler's (1992, 1995) theory of probabilistic sophistication replaces the expected utility operator with a more general non-expected utility function on simple lotteries, but retains the assumption of a prior belief. This approach remains incompatible with ambiguity aversion, i.e., the tendency to prefer objectively defined probabilities (risk) to unknown probabilities (ambiguity), as illustrated by Ellsberg's (1961) classic examples.

To model Ellsberg-type behavior, researchers put forward the notion of a second-order belief—a prior belief on possible distributions of the states—suggesting that the DM makes a probabilistic assessment of the unknown distribution. The best known examples of such models are Segal's (1987) theory of recursive non-expected utility, and the smooth ambiguity model of Klibanoff, Marinacci and Mukerji (2005), which is further studied by Nau (2006), Ergin and Gul (2009), Seo (2009), and Denti and Pomatto (2022), among others. Aside from these specific models, one can also think of a generic form of second-order representations based on a general non-expected utility function on *compound* lotteries, along the lines of Machina and Schmeidler (1992, 1995). In what follows, *second-order probabilistic sophistication* (SOPS) refers to this generic form.

Despite the growing interest in second-order representations, the scope of behavior that can be modeled in this way is not yet fully understood. Notably, Ergin and Gul (2009) show that in a purely subjective setup with no objective lotteries, SOPS can be characterized by Machina and Schmeidler (1992) type axioms, but they assume an enriched state space with two dimensions (issues).[1] By contrast, in a purely subjective setup with a standard state space and monetary prizes, there exists a second-order representation for any rational DM who prefers more to less (see Evren 2017). While the latter representation theorem is extremely general, it does not pin down a unique second-order belief, nor does it allow one to test/compare different versions of the

[1] Following Arrow (1984, p.173), a *state* can be defined as a complete description of the world that identifies the consequences of all relevant actions. The second "issue" in Ergin and Gul's (2009) setup corresponds to the true distribution of such states. For example, in a typical Ellsberg-type experiment, the second issue becomes the (unknown) distribution of colored balls in an urn. Accordingly, Ergin and Gul's axioms refer to (second-order) acts that map first-order distributions to conseqences/prizes. As also noted by Klibanoff et al. (2005), observing preferences on second-order acts is a difficult task, which is a central theme in the discussion between Epstein (2010) and Klibanoff et al. (2012).

theory based on specific assumptions on risk preferences.

In this paper, I study a testable model of SOPS based on two sets of additional consistency assumptions. First, I assume that compound lotteries with objectively defined probabilities are also present in the choice environment, and the DM's ranking of uncertain prospects (acts) is consistent with her actual preferences on those lotteries. In practice, this rules out unrealistic specifications of utility functions on compound lotteries and allows one to compare the descriptive power of different assumptions on risk preferences.

My second set of assumptions are concerned with the description of ambiguity: The DM's second-order belief must be consistent with available information about the distribution of states, and she must update it in a Bayesian fashion when she gets new information about the true distribution. For example, in an Ellsberg-type experiment with colored balls, new information may come in the following form: *The fraction of red balls is at least 30%.* In my representation, the DM responds to this information by conditioning her second-order belief to the set of possible first-order distributions that attach at least 30% probability to the event of extracting a red ball.

The main finding of this paper is a behavioral characterization of the class of Bayesian second-order representations outlined above. First, I formulate some basic axioms that relate the value of an act h conditional on a set Π of first-order distributions to those of simple lotteries induced by h and the distributions in Π. The hypothesis of Bayesian updating entails strong links between conditional beliefs. Using these links, I construct an algorithm that leads to a candidate second-order belief conditional on any given set Π. The algorithm requires the analyst to solve $|\Pi| - 1$ behavioral indifference equations, sequentially, each with one unknown variable. My main axiom is a Bayesian consistency property that focuses on algorithmically constructed candidate conditional beliefs.

More specific versions of the theory are easily obtained by additional, well-known axioms on preferences on compound lotteries. In particular, Segal's (1987) model is characterized by a time-neutrality property, whereas Seo's (2009) version of the smooth ambiguity model demands von Neumann-Morgenstern (vN-M) independence axioms. One can also think of a plethora of alternative specifications that belong to the general class characterized in my main representation. To illustrate potential benefits of this generality, I provide a detailed analysis of a rank-dependent specification that has a number of advantages over both the smooth ambiguity model and Segal's model.

Laboratory experiments (e.g., Halevy 2007, Abdellaoui, Klibanoff and Placido 2015, Chew, Miao and Zhong 2017, and Dean and Ortoleva 2019) document robust connections between attitudes towards ambiguity and compound risk. While these findings provide indirect support for second-order representations, they cannot be taken as conclusive evidence absent a direct test of SOPS.[2] Moreover, in principle, SOPS may correlate with other behavioral patterns such as time neutrality or vN-M independence. Thus, the present model provides a testing ground that may help us better understand the descriptive power of various second-order representations.

In my model, the DM's conditional preference relations uniquely determine her prior second-order belief. This starkly contrasts more standard representations that only focus on ex-ante behavior. For example, within Anscombe and Aumann's (1963) classical expected utility model, two second-order beliefs generate the same behavior whenever they have the same first moment (mean). Seo (2009) notes that this non-uniqueness problem persists in the smooth ambiguity model, aside from some special cases. By extension, in my model, it is also possible to represent the same ex-ante behavior with two different second-order beliefs, but such beliefs tend to produce different reactions to news concerning the true of distribution of the states. That is, different prior beliefs produce different conditional behavior, even if they are indistinguishable ex-ante. This allows me to pin down the DM's prior second-order belief, using the algorithmic approach mentioned above.

On a related note, under fairly general assumptions that are compatible with ambiguity aversion, I show that second-order beliefs with larger spread tend to produce stronger reactions to new information that negatively affects the value of an act. This illustrates how the uniqueness part of my representation may manifest itself in comparative statics exercises.

Halevy and Ozdenoren (2022) study SOPS in a more standard setup with no updating, and highlight the role of the reduction of compound lotteries axiom as a bridge between first and second order probabilistic sophistication. Following Sarin and Wakker's (1997) characterization of the expected utility model, they focus on a calibration axiom that demands the existence of a first-order distribution and a collection of compound lotteries—specifically, two compound lotteries for each pair of disjoint events—that si-

[2] As a partial exception, Chew et al. (2017) perform a test of second-order probabilistic sophistication with two states, two possible distributions and further symmetry assumptions. However, to assess the descriptive performance of second-order representations, they also consider more complicated choice problems that involve more than two possible distributions. My theory also covers such cases.

multaneously solve a number of behavioral (indifference) equations between compound lotteries and acts. If the DM holds a second-order belief, the probabilities that she attaches to possible first-order distributions give a natural solution to the equation system considered in the calibration axiom. However, the equations in the axiom may have multiple solutions, and in those cases, the axiom does not pin down a representing second-order belief. This identification problem is parallel to the aforementioned non-uniqueness issue, and seems to further illustrate the contribution of my identification algorithm, which becomes functional under the Bayesian updating hypothesis.

The second-order Bayesian representation is formally introduced in Section 2 and characterized in Section 3. I discuss special versions of the theory in Section 4, while Section 5 focuses on comparative statics of second-order beliefs. Section 6 highlights some limitations of second-order representations in general, and my Bayesian approach in particular. The proofs of the main results are in the appendix.

2. Setup

Let A be an arbitrary set. The set of all probability measures on A (with finite support) is denoted by $\Delta(A)$. As usual, $\eta(B)$ stands for the probability of a set $B \subseteq A$ according to $\eta \in \Delta(A)$. I write $\eta(a)$ in place of $\eta(\{a\})$. By assumption, $\text{supp}(\eta) := \{a \in A : \eta(a) > 0\}$ is a finite set, and $\sum_{a\in A} \eta(a) = 1$. Given $\{\eta^1, ..., \eta^n\} \subseteq \Delta(A)$ and $\{\alpha^1, ..., \alpha^n\} \subseteq [0, 1]$ with $\sum_{i=1}^n \alpha^i = 1$, the mixture $\sum_{i=1}^n \alpha^i \eta^i$ is the element of $\Delta(A)$ that attaches the probability $\sum_{i=1}^n \alpha^i \eta^i(a)$ to $a \in A$. Similarly, $\Delta^2(A) := \Delta(\Delta(A))$ is the set of all probability measures on $\Delta(A)$, and the mixture operation on this set is defined analogously. Given a subset B of A, I do not distinguish between the elements of $\Delta(B)$ and those elements η of $\Delta(A)$ such that $\text{supp}(\eta) \subseteq B$.

The degenerate probability measure supported on a point $a \in A$ is denoted as δ_a. For any $B \subseteq A$, we have $\delta_a(B) := 1$ if $a \in B$, and $\delta_a(B) := 0$ otherwise. Note that $\eta = \sum_{a\in A} \eta(a)\delta_a$ for every $\eta \in \Delta(A)$.

The set of prizes is an interval $X \subseteq \mathbb{R}$ with $\inf X < \sup X$. A generic element of $\Delta(X)$, denoted q (or $q', \hat{q}$ etc.), corresponds to a simple (or one-shot) lottery on X. In turn, a generic element of $\Delta^2(X)$, denoted Q, represents a compound lottery. The sets $\Delta(X)$ and $\Delta^2(X)$ are equipped with the topology of weak convergence.[3] I often write $\Delta(X)$ in place of the set $\{\delta_q : q \in \Delta(X)\} \subseteq \Delta^2(X)$, and X in place of

[3] A sequence (q^n) in $\Delta(X)$ converges to $q \in \Delta(X)$ iff $\sum_{x\in X} q^n(x)f(x) \to \sum_{x\in X} q(x)f(x)$ for every continuous and bounded $f : X \to \mathbb{R}$. The convergence criterion on $\Delta^2(X)$ is analogous.

$\{\delta_x : x \in X\} \subseteq \Delta(X)$.

The first-order stochastic dominance relation on $\Delta(\mathbb{R})$ and its strict part are denoted by $\geq_{fsd}$ and $>_{fsd}$, respectively. Given an interval $Z \subseteq \mathbb{R}$, I say that a function $\varphi : \Delta(Z) \to \mathbb{R}$ is *fsd-increasing* if $q >_{fsd} q'$ implies $\varphi(q) > \varphi(q')$ for every $q, q' \in \Delta(Z)$. The DM's preferences on compound lotteries will be represented[4] by a function $V : \Delta^2(X) \to \mathbb{R}$ that is *twice fsd-increasing* in the following sense: (i) $v(q) \equiv V(\delta_q)$ is fsd-increasing on $\Delta(X)$; and (ii) $\sum_{q\in\Delta(X)} Q(q)\delta_{v(q)} \geq_{fsd} (>_{fsd}) \sum_{q\in\Delta(X)} Q'(q)\delta_{v(q)}$ implies $V(Q) \geq (>) V(Q')$, meaning that V is monotonically increasing with respect to $\geq_{fsd}$ on simple utility distributions induced by compound lotteries. Parts (iii) and (iv) of Axiom (A5) in Section 3 provide a behavioral description of such representations.

Lemma 1. *Suppose $V : \Delta^2(X) \to \mathbb{R}$ is continuous and twice fsd-increasing. Then there exists a continuous and fsd-increasing function $\Phi : \Delta(v(X)) \to \mathbb{R}$ such that $V(Q) = \Phi(\sum_{q\in\Delta(X)} Q(q)\delta_{v(q)})$ for every $Q \in \Delta^2(X)$. Moreover, given any continuous function $u : \Delta(X) \to \mathbb{R}$ that is ordinally equivalent to v, so that $u(q) \geq u(q')$ iff $v(q) \geq v(q')$ for every $q, q' \in \Delta(X)$, there exists a continuous and increasing function $\phi : u(X) \to \mathbb{R}$ such that $v(q) = \phi(u(q))$ for every $q \in \Delta(X)$.*

Thus, a twice fsd-increasing function V can equivalently be thought of as a triplet (Φ, ϕ, u), where u is an arbitrarily selected function that represents the DM's preferences on simple lotteries; ϕ is an increasing transformation that solves the functional equation $v = \phi \circ u$; and Φ is an aggregator that determines the overall utility of a compound lottery Q as a function of the utility distribution in $\Delta(\phi \circ u(X))$ induced by Q. I refer to the triplet (Φ, ϕ, u) as a *simple-form representation* because it is based on simple distributions on utility levels. In Section 4, I discuss particular variants of such representations associated with some well-known models. (For brevity, I omit the proof of Lemma 1, which is a straightforward exercise.)

Let S be a finite set that consists of the states of nature. *A first-order distribution* refers to a generic element of $\Delta(S)$, denoted by p. Similarly, a *second-order distribution* is a generic element P of $\Delta^2(S)$. A function that maps S into X is referred to as an *act*. The set of all acts is $\mathcal{H} := X^S$, with generic elements h and g.

If the underlying uncertainty can accurately be described by a first-order distribu-

[4]Given a binary relation $\succsim$ on a set A and a subset $B \subseteq A$, I say that a function $U : B \to R$ *represents $\succsim$ on B* provided that $U(a) \geq U(a')$ iff $a \succsim a'$ for every $a, a' \in B$.

tion $p \in \Delta(S)$, an act h induces the following simple lottery:

$$p_h := \sum_{s \in S} p(s)\delta_{h(s)}.$$

In choice problems under ambiguity, the DM may not be able to pick out such a distribution. Suppose that she has in mind a second-order distribution/belief $P \in \Delta^2(S)$, where $P(p)$ is her assessment of the likelihood of the event that the states are distributed according to p. Then an act h induces a compound lottery:

$$P_h := \sum_{p \in \Delta(S)} P(p)\delta_{p_h}.$$

Definition 1. A binary relation $\succsim'$ on $\mathcal{H} \cup \Delta^2(X)$ is *second-order probabilistically sophisticated* (SOPS) if there exist a continuous, twice fsd-increasing function $V : \Delta^2(X) \to \mathbb{R}$ and a second-order distribution $P \in \Delta^2(S)$ such that for any $\rho, \hat{\rho} \in \mathcal{H} \cup \Delta^2(X)$,

$$\rho \succsim' \hat{\rho} \quad \Leftrightarrow \quad \mathcal{V}_P(\rho) \geq \mathcal{V}_P(\hat{\rho}),$$

where $\mathcal{V}_P(h) := V(P_h)$ for $h \in \mathcal{H}$, and $\mathcal{V}_P(Q) := V(Q)$ for $Q \in \Delta^2(X)$. I denote by $\succsim_{(V,P)}$ the SOPS relation $\succsim'$ on $\mathcal{H} \cup \Delta^2(X)$ that is represented by a given pair (V, P) in this fashion.

As a key feature, an SOPS relation $\succsim'$ satisfies $h \sim' P_h$ for every $h \in \mathcal{H}$. That is, the DM is indifferent between an act h and the compound lottery induced by h and her second-order belief P.[5]

I assume that given the initial information available to the DM, the true distribution of the states belongs to an exogenously given, finite set $\Pi^* \subseteq \Delta(S)$. Nonempty subsets of Π^* are denoted as Π, Π' etc., while 2^{Π} stands for the collection of all nonempty subsets of Π.

At an interim stage, the DM learns that the true distribution of the states lies in a set Π. Her behavior upon receiving this new information is described by a binary relation $\succsim_\Pi$ on $\mathcal{H} \cup \Delta^2(X)$. Accordingly, the primitive of the model is a collection of binary relations $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ on the set $\mathcal{H} \cup \Delta^2(X)$. I often write $\succsim$ in place of the ex-ante preference $\succsim_{\Pi^*}$.

I proceed with an example that illustrates the setup.

[5] I denote by $\sim$ and $\succ$ the symmetric and asymmetric parts of a relation $\succsim$, respectively.

Example 1. Suppose that the DM is asked to bet on the color of a ball to be randomly extracted from an urn that contains nine balls. Initially, she only knows that three of the balls are red (r), while the remaining balls are either black (b) or white (w). Then the state space is $S := \{r, b, w\}$, where each state represents the event of extracting a ball of the corresponding color. Given the set of possible color distributions in the urn, we can let $\Pi^* := \{p \in \Delta(S) : p(r) = 1/3, p(b) = \alpha/9, \alpha = 0, 1, ..., 6\}$. Before choosing a bet, the DM gets a further piece of information on the color distribution in the urn. For instance, if she learns that there are at most two black balls, the set of relevant distributions becomes $\Pi := \{p \in \Pi^* : p(b) \leq 2/9\}$.

My main purpose is to characterize, behaviorally, the representation notion defined below.

Definition 2. A *second-order Bayesian representation* (V, P^*) for the collection of preferences $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ consists of a continuous, twice fsd-increasing function $V : \Delta^2(X) \to \mathbb{R}$ and a distribution $P^* \in \Delta(\Pi^*)$ such that $\succsim_\Pi = \succsim_{(V,P^*|\Pi)}$ for every $\Pi \in 2^{\Pi^*}$ with $P^*(\Pi) > 0$. Here, $P^*|\Pi$ is the Bayesian update defined as $P^*|\Pi(p) := P^*(p)/P^*(\Pi)$ for $p \in \Pi$, and $P^*|\Pi(p) := 0$ for $p \in \Pi^* \backslash \Pi$.

The representation describes a Bayesian DM who holds an ex-ante second-order belief $P^* \in \Delta(\Pi^*)$. Upon learning that the true distribution lies in a set Π with $P^*(\Pi) > 0$, she updates P^* according to the Bayes rule, converts acts into compound lotteries via the updated second-order belief, $P^*|\Pi$, and evaluates those lotteries with the function V. Shortly put, $\succsim_\Pi$ is the SOPS relation represented by the pair $(V, P^*|\Pi)$.

3. Representation Theorem

In the remainder of the paper, I often write p_h in place of δ_{p_h}, and more generally, q in place of δ_q. I say that a set $\Pi \in 2^{\Pi^*}$ is *null* if for each $p \in \Pi$ there exist a $p' \in \Pi^*$ and an $h \in \mathcal{H}$ such that $p'_h \succ p_h$ and $h \sim_{\{p,p'\}} p'_h$. The latter two conditions suggest that information concerning the (null) distribution p does not influence the DM's behavior. After learning that the true distribution is either p or p', her evaluation of the act h solely depends on p'.

My first axiom demands a similar form of indifference toward all null sets: Learning that the true distribution does not belong to a null-set Π should not influence the DM's behavior.

A1: Null Irrelevance (NI). (i) If Π is null and distinct from Π^*, then $\succsim_{\Pi^*\backslash\Pi} =$

$\succsim_{\Pi^*}$.

(ii) If Π is null and Π' is non-null, then $\succsim_{\Pi'} = \succsim_{\Pi' \cup \Pi}$.

Part (ii) of this axiom presupposes the existence of a non-null set Π'. If we were to assume that the grand set Π^* is non-null, statement (i) would immediately follow from statement (ii). Indeed, the only role of NI(i) in my analysis is to ensure that the set Π^* is non-null.

A2: Singleton Sophistication (SiSo). If $\{p\}$ is non-null, then $h \sim_{\{p\}} p_h$ for every $h \in \mathcal{H}$.

As a minimal sophistication condition, our DM must be able to calculate the lottery p_h induced by an act h and a distribution p, and evaluate h in this manner if she learns that p is the true distribution of the states. This is the content of SiSo.

A3: Monotonic Expansion (ME). If Π is non-null and $p'_h \succsim p_h \succsim p''_h$ for every $p \in \Pi$, then $h \sim_\Pi Q$ implies $h \succsim_{\Pi \cup \{p'\}} Q$ and $Q \succsim_{\Pi \cup \{p''\}} h$.

Suppose $p'_h \succsim p_h$ for every $p \in \Pi$, where p' is a distribution that does not belong to Π. Let us consider two different scenarios regarding the interim stage. In one scenario, the DM learns that the true distribution belongs to Π, and in the second scenario she learns that p' can also be true in addition to those distributions in Π. ME asserts that h should be more valuable in the second scenario than the first, measured against the values of lotteries $Q \in \Delta^2(X)$. Analogously, the converse conclusion must hold whenever $p_h \succsim p'_h$ for every $p \in \Pi$.

If we think of $\Pi \cup \{p'\}$ and Π as nested events unfolding sequentially, and take p_h as the outcome of the act h under a distribution p, ME amounts to saying that the value of h can only decrease (resp. increase) for the DM once she learns that the best (resp. worst) possible outcome that she can expect from h is not actually feasible.

Assuming a Bayesian DM who holds a second-order belief, variations in the DM's behavior conditional on nested sets of distributions can be used to elicit her assessment of the distributions' relative likelihood. The next definition introduces an algorithm that determines a second-order distribution on any given non-null event Π by iteratively extending pre-determined distributions supported on subsets. Following the logic of ME, the algorithm focuses on a single, outstanding distribution at each step of the iteration.

Definition 3. For any non-null $\Pi \in 2^{\Pi^*}$, a distribution $P \in \Delta(\Pi)$ is *pre-consistent on* Π (with respect to the collection of conditional preferences) if one of the two statements below holds true:

1. $\Pi = \{p\}$ for some $p \in \Pi^*$.

2. There exist a $p' \in \Pi$, an act g with $p'_g \succ p_g$ for every $p \in \Pi \backslash \{p'\}$, a distribution $\hat{P} \in \Delta(\Pi \backslash \{p'\})$ that is pre-consistent on $\Pi \backslash \{p'\}$, and a number $\gamma \in [0,1]$ such that

$$g \sim_\Pi \gamma \delta_{p'_g} + (1-\gamma)\hat{P}_g \quad \text{and} \quad P = \gamma \delta_{p'} + (1-\gamma)\hat{P}.$$

The first part of Definition 3 declares the degenerate distribution δ_p pre-consistent in the unambiguous case $\Pi = \{p\}$. The second part of the definition describes how a pre-consistent distribution $\hat{P}$ on an event $\hat{\Pi}$ can be extended to an event $\Pi := \hat{\Pi} \cup \{p'\}$. First of all, there must exist an act g such that $p'_g \succ p_g$ for every $p \in \hat{\Pi}$. This ensures that the utility of the mixture $\gamma \delta_{p'_g} + (1-\gamma)\hat{P}_g$ is increasing in γ, assuming a twice fsd-increasing utility function. If g is indifferent to $\gamma \delta_{p'_g} + (1-\gamma)\hat{P}_g$ conditional on Π for a particular value of γ, then the corresponding mixture $P = \gamma \delta_{p'} + (1-\gamma)\hat{P}$ is declared pre-consistent on Π. Indeed, $\gamma \delta_{p'} + (1-\gamma)\hat{P}$ is the unique distribution on Π that assigns probability γ to p' while agreeing with $\hat{P}$ conditional on $\hat{\Pi}$, where γ is the "correct" weight of p' suggested by the behavioral equation $g \sim_\Pi \gamma \delta_{p'_g} + (1-\gamma)\hat{P}_g$.

A distribution $P \in \Delta(\Pi)$ constructed via the algorithm in Definition 3 is called "pre-consistent" because a further consistency condition is necessary to guarantee the existence of a second order Bayesian representation.

A4: Consistent Expansion (CE). Let $P \in \Delta(\Pi)$ be a pre-consistent distribution on a non-null set Π. Given any non-null $\{p^1, p^2\} \subseteq \Pi$, $\alpha \in [0,1]$ and $h \in \mathcal{H}$, set $Q := \alpha \delta_{p^1_h} + (1-\alpha)\delta_{p^2_h}$. Then we have

$$h \succsim_{\{p^1,p^2\}} Q \quad \Leftrightarrow \quad h \succsim_\Pi P(\{p^1,p^2\})Q + \textstyle\sum_{p \in \Pi \backslash \{p^1,p^2\}} P(p)\delta_{p_h}.$$

Suppose that the algorithm that we just discussed leads to a second-order distribution P supported on a set Π. Roughly speaking, CE demands this distribution be consistent with preferences conditional on binary subsets of Π. The ranking of an act h relative to the lottery $Q := \alpha \delta_{p^1_h} + (1-\alpha)\delta_{p^2_h}$ conditional on a subset $\{p^1, p^2\}$ must be the same as that of h relative to the linear transformation $P(\{p^1,p^2\})Q + \sum_{p \in \Pi \backslash \{p^1,p^2\}} P(p)\delta_{p_h}$ conditional on the large set Π. Alternatively, CE can be viewed as an additive separability property based on carefully selected weights embodied in a pre-consistent distribution. Section 3.1 contains further remarks on pre-consistent distributions and the implications of CE.

My final axiom consists of some standard properties.

A5: Standard Properties (STD). (i) For any non-null Π, the relation $\succsim_\Pi$ on $\mathcal{H} \cup \Delta^2(X)$ is complete, transitive, and continuous in the sense that $\{Q \in \Delta^2(X) : Q \succ \rho\}$ and $\{Q \in \Delta^2(X) : \rho \succ Q\}$ are open subsets of $\Delta^2(X)$ for each $\rho \in \mathcal{H} \cup \Delta^2(X)$.
(ii) For any non-null Π and $Q, Q' \in \Delta^2(X)$, we have $Q \succsim_\Pi Q'$ if and only if $Q \succsim Q'$.
(iii) For any $q, q' \in \Delta(X)$ with $q >_{fsd} q'$, we have $q \succ q'$.
(iv) For any $\{\hat{q}^0, q^0, q^1, ..., q^n\} \subseteq \Delta(X)$ and $\{\alpha^0, \alpha^1, ..., \alpha^n\} \subseteq (0,1)$ with $\sum_{i=0}^n \alpha^i = 1$, we have $\alpha^0 \delta_{q^0} + \sum_{i=1}^n \alpha^i \delta_{q^i} \succsim \alpha^0 \delta_{\hat{q}^0} + \sum_{i=1}^n \alpha^i \delta_{q^i}$ iff $q^0 \succsim \hat{q}^0$.

Statements (iii) and (iv) above correspond to the monotonicity properties of a twice fsd-increasing function on $\Delta^2(X)$. Property (iv) is also known as the "compound independence axiom" (see Segal 1990). Statement (ii) means that the DM's risk preferences do not to vary with new information on possible distributions of the states. This is a standard assumption in the updating literature, and so are the remaining assumptions in statement (i).

To avoid trivialities, I assume that the collection of conditional preferences is *non-constant* in the sense that $\succsim_\Pi \neq \succsim_{\Pi'}$ for some $\Pi, \Pi' \in 2^{\Pi^*}$. The following is the main result of the paper.

Theorem 1. *A non-constant collection of binary relations* $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ *on* $\mathcal{H} \cup \Delta^2(X)$ *satisfies the axioms* (A1)-(A5) *if and only if it admits a second-order Bayesian representation* (V, P^*). *Moreover,* P^* *is uniquely defined.*

This theorem provides a characterization of second-order Bayesian behavior by means of testable axioms. The axioms link the conditional values of a subjective act h to those of compound lotteries jointly induced by h and possible distributions of the states. Aside from two basic monotonicity properties, my axioms do not restrict the DM's preferences on compound lotteries. Compatible forms of ambiguity attitudes and preferences on subjective acts are also unlimited, except for necessary connections between the valuations of acts and compound lotteries that underlie second-order Bayesian representations.

In Section 4, we will see that interesting forms of the function V can easily be characterized by additional axioms on preferences on compound lotteries. Thus, Theorem 1 provides a unifying framework that can be used to test and compare the descriptive power of various second-order representations within a population of subjects who update their second-order belief in a Bayesian fashion.

3.1. Further Remarks

In the second part of Definition 3, we may as well assume that the distribution $\hat{P} \in \Delta(\Pi \backslash \{p'\})$ satisfies

$$h \sim (\hat{P}|\Pi')_h \quad \forall h \in \mathcal{H} \text{ and non-null } \Pi' \in 2^{\Pi\backslash\{p'\}}. \tag{1}$$

Then a pre-consistent distribution P on Π becomes a natural, one-step extension of a second-order belief, $\hat{P}$, that is known to represent preferences conditional on subsets of $\Pi\backslash\{p'\}$, while CE reduces to a statement on extensibility of such representations. Yet, in practice, condition (1) may complicate the task of eliciting pre-consistent beliefs from choice data. I have therefore chosen to focus on the recursive approach in Definition 3.

Upon letting $p^1 = p^2$ in the statement of CE, from SiSo and CE we get $h \sim_\Pi P_h$ for every $h \in \mathcal{H}$. Thus, once we elicit a pre-consistent distribution on a given set Π, the conditional preference $\succsim_\Pi$ must be SOPS as an immediate consequence of the axioms. In turn, with $p^1 \neq p^2$, CE helps establish Bayesian relations between pre-consistent distributions on nested subsets of Π^*.

A second-order distribution that represents a SOPS preference relation is not necessarily unique. This is especially commonplace for second-order formulations of Anscombe and Aumann's (1963) classical expected utility model. The next example illustrates how Bayesian updating achieves uniqueness in Theorem 1.

Example 2. Let $\Pi^* := \{p^1, p^2, p^3\}$, where $p^1 \neq p^3$ and $p^2 = \frac{1}{2}p^1 + \frac{1}{2}p^3$. Pick an fsd-increasing expected utility function $u : \Delta(X) \to \mathbb{R}$. Set $V(Q) := \sum_{q\in Q} Q(q)u(q)$ for $Q \in \Delta^2(X)$, and $P^\lambda := \lambda\delta_{p^2} + (1-\lambda)(\frac{1}{2}\delta_{p^1} + \frac{1}{2}\delta_{p^3})$ for $\lambda \in (0,1)$. It is easily checked that $V(P^\lambda_h) = u\,(p^2_h)$ for any $h \in \mathcal{H}$, independently from the choice of λ, because both V and u are expected utility functions. But if $u(p^1_h) \neq u(p^3_h)$—such h exists because u is fsd-increasing, then $V((P^\lambda|\{p^1,p^2\})_h) = u(\frac{1-\lambda}{1+\lambda}p^1_h + \frac{2\lambda}{1+\lambda}p^2_h)$ returns distinct values as a function of λ. Thus, the choice of λ influences the conditional preference $\succsim_{\{p^1,p^2\}}$ associated with P^λ but not the corresponding ex-ante preference.

In models of first-order Bayesian updating, null sets are defined by an analogue of NI(i) formulated in terms of step functions on S. In the present setup, this approach is not so suitable because a non-null Π can also satisfy the conclusion of NI(i). That is, unlike the first-order Bayesian theory, learning an event Π does not necessarily alter the DM's preferences even if both Π and its complement $\Pi^*\backslash\Pi$ are non-null. For instance, in Example 2, $P^\lambda|\{p^2\}$, $P^\lambda|\{p^1,p^3\}$ and P^λ induce the same preference on $\mathcal{H}$, for any given $\lambda \in (0,1)$. Yet P^λ attaches positive probabilities to all three distributions, p^1, p^2 and p^3. This manifests itself through the pattern $p^i_h \succ_{\{p^i,p^j\}} h \succ_{\{p^i,p^j\}} p^j_h$ for any act h with $p^i_h \succ p^j_h$.

Example 2 also shows that in the second part of Definition 3, it is important to focus on an outstanding distribution p' such that $p'_g \succ p_g$ for every $p \in \Pi \backslash \{p'\}$. To meet this condition, in the context of Example 2 one can start from the intermediate distribution $\{p^2\}$, then identify a pre-consistent distribution on a doubleton $\{p^2, p^i\}$, which can be extended further to a pre-consistent distribution on the grand set $\Pi^* = \{p^1, p^2, p^3\}$. By contrast, it would be futile to attempt to reach the ex-ante belief P^λ from the conditional belief on the set $\{p^1, p^3\}$. Indeed, with $\hat{P} := P^\lambda | \{p^1, p^3\} = \frac{1}{2}\delta_{p^1} + \frac{1}{2}\delta_{p^3}$ and $p' := p^2$, the equation $g \sim_{\Pi^*} \gamma \delta_{p'_g} + (1-\gamma)\hat{P}_g$ is completely uninformative about the likelihood of p'; it holds for any $g \in \mathcal{H}$ and $\gamma \in [0,1]$.

4. Special Forms

The function V in a second-order Bayesian representation determines the DM's attitudes toward compound risk and ambiguity simultaneously. In this section, I discuss two well-known specifications of this function, and an alternative that accounts for some recent empirical findings.

Let (Φ, ϕ, u) denote a simple-form expression of a continuous, twice fsd-increasing function $V : \Delta^2(X) \to \mathbb{R}$ that represents the DM's preferences on compound lotteries, $\succsim$. By definitions, $V(Q) = \Phi(\sum_{q \in \Delta(X)} Q(q)\delta_{v(q)})$ for every $Q \in \Delta^2(X)$, and $v = \phi \circ u$.

In the smooth ambiguity model, both u and Φ are expected utility operators, with $u(q) = E_q(u) := \sum_{x \in X} q(x)u(x)$ for $q \in \Delta(X)$, and $\Phi(\chi) = \sum_{w \in v(X)} \chi(w)w$ for $\chi \in \Delta(v(X))$. Thus, the function V has the following form:

$$V_{Sm}(Q) := \sum_{q \in \Delta(X)} Q(q)\phi\left(E_q(u)\right).$$

Segal's (1987) model builds upon a non-expected utility function u on simple lotteries; the rank-dependent utility model (Quiggin 1982) and the cautious expected utility model (Cerreia-Vioglio, Dillenberger and Ortoleva 2015) are the most popular specifications for this function. There are two other key features of Segal's model. First, u and v are normalized so that $u(q) = v(q) = x(q)$ for each $q \in \Delta(X)$, where $x(q) \in X$ is the certainty equivalent of q with $\delta_{x(q)} \sim q$ (or, more precisely, $\delta_{\delta_{x(q)}} \sim \delta_q$). This implies $v(X) = X$. Second, $\Phi(q) = u(q)$ for every $q \in \Delta(X)$, meaning that the function $u(\cdot) = x(\cdot)$ is applied two times, recursively, to evaluate a compound lottery. To summarize, the utility of $Q \in \Delta^2(X)$ is defined as follows:

$$V_{Se}(Q) := x\left(\sum_{q \in \Delta(X)} Q(q)\delta_{x(q)}\right).$$

An alternative specification utilizes a rank-dependent aggregator Φ in place of the first-stage expected utility operator of the smooth ambiguity model. The utility of a compound lottery $Q = \sum_{i=1}^{n} Q(q^i)\delta_{q^i} \in \Delta^2(X)$ with $q^n \succsim \cdots \succsim q^1$ is given by

$$V_{Sr}(Q) := \phi(E_{q^1}(u)) + \sum_{j=2}^{n} \left(\phi(E_{q^j}(u)) - \phi(E_{q^{j-1}}(u))\right) \Psi(\textstyle\sum_{i=j}^{n} Q(q^i)).$$

Here, $\Psi : [0,1] \to [0,1]$ is a continuous and increasing probability transformation function with $\Psi(0) = 0$ and $\Psi(1) = 1$, while the functions u and ϕ are defined as in the smooth ambiguity model. In what follows, I refer to this specification as SORDU (the second-order rank-dependent utility model).

As we shall see shortly, the second-order Bayesian representations corresponding to these three specifications are easily characterized by some well-known assumptions on risk preferences.

On a more general note, it is possible to combine any specification of $u : \Delta(X) \to \mathbb{R}$ with any specification of $\Phi : \Delta(v(X)) \to \mathbb{R}$ that may be of interest. Naturally, the structure of u is determined by preferences on the set $\{\delta_q : q \in \Delta(X)\}$, which we do not distinguish from $\Delta(X)$. In turn, preferences on the set $\Delta_o := \{Q_o \in \Delta^2(X) : \mathrm{supp}(Q_o) \subseteq \{\delta_x : x \in X\}\}$, the set of compound lotteries that are degenerate in the second stage, determine the structure of the first-stage aggregator Φ. Note that the rule $q_o(x) := Q_o(\delta_x)$ establishes an isomorphism between $\Delta(X)$ and Δ_o. Thus, familiar forms of Φ can also be characterized by known axioms on preferences on simple lotteries.

Proposition 1. *Let $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ be a non-constant collection of binary relations on $\mathcal{H} \cup \Delta^2(X)$ that satisfies the axioms* (A1)-(A5). *This collection admits a second-order Bayesian representation (V, P^*) with:*
(i) *$V = V_{Sm}$ iff $\succsim$ satisfies* (a) *first-stage vN-M independence: $Q \succsim Q'$ implies $\alpha Q + (1-\alpha)\hat{Q} \succsim \alpha Q' + (1-\alpha)\hat{Q}$ for $\alpha \in (0,1)$ and $Q, Q', \hat{Q} \in \Delta^2(X)$; and* (b) *second-stage vN-M independence: $\delta_q \succsim \delta_{q'}$ implies $\delta_{\alpha q + (1-\alpha)\hat{q}} \succsim \delta_{\alpha q' + (1-\alpha)\hat{q}}$ for $\alpha \in (0,1)$ and $q, q', \hat{q} \in \Delta(X)$.*
(ii) *$V = V_{Se}$ iff $\succsim$ satisfies time neutrality: $Q_o \sim \delta_{q_o}$ for $Q_o \in \Delta_o$, where $q_o \in \Delta(X)$*

is defined as $q_o(x) := Q_o(\delta_x)$ *for* $x \in X$.

(iii) $V = V_{Sr}$ *iff* $\succsim$ *satisfies second-stage vN-M independence, and first-stage weak commutativity: Let* $q, \hat{q} \in \Delta(X)$ *and* $Q, \hat{Q} \in \Delta^2(X)$ *be such that* $\delta_q \sim Q = \sum_{i=1}^n \alpha^i \delta_{q^i}$ *and* $\delta_{\hat{q}} \sim \hat{Q} = \sum_{i=1}^n \alpha^i \delta_{\hat{q}^i}$ *for some* $\{q^1, ..., q^n, \hat{q}^1, ..., \hat{q}^n\} \subseteq \Delta(X)$ *and* $\{\alpha^1, ..., \alpha^n\} \subseteq (0,1)$ *with* $\sum_{i=1}^n \alpha^i = 1$. *Assume further that* $\delta_{q^n} \succsim \cdots \succsim \delta_{q^1}$, $\delta_{\hat{q}^n} \succsim \cdots \succsim \delta_{\hat{q}^1}$, *and* $\delta_{q^n} \succsim \delta_{\hat{q}^n}, ..., \delta_{q^1} \succsim \delta_{\hat{q}^1}$. *Then for any* $\lambda \in (0,1)$ *and* $\{\bar{q}^1, ..., \bar{q}^n\} \subseteq \Delta(X)$ *with* $\delta_{\bar{q}^i} \sim \lambda\delta_{q^i} + (1-\lambda)\delta_{\hat{q}^i}$ *for* $i = 1, ..., n$, *we have* $\lambda\delta_q + (1-\lambda)\delta_{\hat{q}} \sim \sum_{i=1}^n \alpha^i \delta_{\bar{q}^i}$.

In the characterization of SORDU, Chew's (1989) weak commutativity axiom is applied on compound lotteries.[6] First-stage vN-M independence takes the role of weak commutativity in the characterization of the smooth ambiguity model. Finally, Segal's model is characterized by the time neutrality property, which asserts that each compound lottery Q_o in Δ_o must be indifferent to the corresponding simple lottery q_o. Needless to say, one can add further axioms on preferences on simple lotteries to characterize more structured versions of Segal's model with a specific family of the non-expected utility function u.

Part (ii) of Proposition 1 is arguably more important than part (i) because Segal's model does not have a compelling characterization in the earlier literature.[7] Although the role of the time neutrality axiom is well-understood, difficulties associated with the characterization of SOPS are responsible from this gap in the literature.

Compared to Segal's model, a key difference of SORDU is the expected utility function $u(q) = E_q(u)$ on simple lotteries, which is likely to be useful for modelers who wish to think of ambiguity aversion/Ellsberg-type behavior as a distinct concept than Allais-type risk preferences. Indeed, most ambiguity models in the literature—including the smooth ambiguity model—employ vN-M preferences on simple lotteries.

In the next section, I show that SORDU overcomes some descriptive shortcomings of the smooth ambiguity model, and has some additional analytical advantages over both alternative theories.

[6] The weak commutativity axiom is also known as "weak certainty equivalent substitution" (see Schmidt 2004). In line with my earlier comments, the proof of Proposition 1 shows that to characterize SORDU it is actually sufficient to confine first-stage weak commutativity to the set Δ_o; that is, without loss of generality, we can assume that all simple lotteries in the statement of the axiom (i.e., $q, \hat{q}, q^i, \hat{q}^i$ and $\bar{q}^i$, $i = 1, ..., n$) are degenerate. One can also utilize Wakker's (1994) tradeoff consistency axiom in place of weak commutativity.

[7] Section 6 provides a discussion of recent developments on the foundations of the smooth ambiguity model.

4.1. Analytical and Descriptive Features of SORDU

Non-participation in financial markets. Empirical findings show that the fraction of individuals who do not participate in equity markets is quite high, even among wealthy people (Vissing-Jorgensen 2003; Briggs, Cesarini, Lindqvist and Östling 2021). Since Dow and Werlang (1992), we know that non-expected utility models—a non-additive prior, in particular—can generate non-participation in financial markets under permissive assumptions on parameter values. In SORDU, the value of an act h, $V_{Sr}(P_h)$, equals the Choquet integral of the function $p \to \phi\left(E_{p_h}(u)\right)$ with respect to the non-additive probability $\Psi \circ P$ on subsets of Π^*. If we take p as the true distribution of returns associated with a financial asset, SORDU leads to similar predictions as in Dow and Werlang, assuming that the second-order distribution P is non-degenerate.

Segal's (1987) model is also compatible with large non-participation rates, but this holds, to a lesser extent, even when P is degenerate (see Evren 2019). Thus, the role of ambiguity in Segal's model is not as sharp.

Ambiguity attitudes vs. the level of ambiguity. The curvature of the function ϕ in the smooth ambiguity model determines ambiguity attitudes. When ϕ is concave, the DM is *absolutely ambiguity averse* in the sense that, where $\mu(P) := \sum_{p \in \Delta(S)} P(p)p$, we have

$$\mu(P)_h \succsim P_h \quad \forall h \in \mathcal{H}. \tag{2}$$

Here, the first-order distribution $\mu(P) \in \Delta(S)$ is the mean or (reduced form) of the second-order distribution $P \in \Delta^2(S)$. Thus, property (2) says that the DM would like the uncertain acts more if she were able to replace her second-order belief P with its mean, $\mu(P)$, and eliminate the ambiguity in the environment.

Similarly, if ϕ is a concave transformation of another function $\hat{\phi}$, the former induces relatively more ambiguity averse behavior in the smooth ambiguity model.[8] Thereby, the smooth ambiguity model allows one to study ambiguity attitudes independently from the level of ambiguity, which can be measured with the spread of a second-order belief P.

By contrast, Evren (2019) shows that such a separation is not possible in Segal's (1987) model, at least if one is willing to maintain a clear distinction between simple-risk

[8] Suppose $\phi = \varphi \circ \hat{\phi}$ for an increasing, concave function $\varphi : \hat{\phi}(u(X)) \to \mathbb{R}$. Let V_{Sm} and $\hat{V}_{Sm}$ denote the smooth ambiguity functions associated with ϕ and $\hat{\phi}$, respectively. Then for any $P \in \Delta^2(S)$, $q \in \Delta(X)$, and $h \in \mathcal{H}$ with $V_{Sm}(P_h) \geq V_{Sm}(\delta_q)$, we also have $\hat{V}_{Sm}(P_h) \geq \hat{V}_{Sm}(\delta_q)$. That is, whenever V_{Sm} ranks h above q, so does $\hat{V}_{Sm}$.

aversion and ambiguity aversion. Indeed, a second-order belief is the only parameter in Segal's model aside from the preference relation on simple lotteries.

In SORDU, the curvature of ϕ influences relative ambiguity attitudes just as in the smooth ambiguity model because $V_{Sr}(P_h)$ equals the expectation of $\phi(E_{p_h}(u))$ with respect to a distorted probability measure on the set Π^*.[9]

The function Ψ in SORDU provides an additional channel that influences ambiguity attitudes. If we replace Ψ with a function $\hat{\Psi}$ such that $\Psi(\alpha) \geq \hat{\Psi}(\alpha)$ for every $\alpha \in (0, 1)$, then the values of acts (and compound lotteries) become smaller, but the values of riskless prizes (and simple lotteries) remain the same. Consequently, the function $\hat{\Psi}$ induces a smaller certainty equivalent for each act, implying an increase in ambiguity aversion.

Next, I highlight some peculiar features of SORDU concerning absolute ambiguity aversion.

Absolute ambiguity aversion vs. compound-risk aversion. In SORDU, concavity of the function ϕ is neither necessary nor sufficient for absolute ambiguity aversion. Rather, concavity of ϕ and convexity of Ψ jointly characterize a *strong form of compound-risk aversion*: Where $\mu(Q) := \sum_{q \in \Delta(X)} Q(q)q$,

$$\alpha\hat{Q} + (1-\alpha)\delta_{\mu(Q)} \succsim \alpha\hat{Q} + (1-\alpha)Q \quad \forall Q, \hat{Q} \in \Delta^2(X) \text{ and } \alpha \in [0,1]. \tag{3}$$

Since $\mu(Q)$ is the mean of the compound lottery Q, this property describes a DM who dislikes a mean-preserving spread in compound risk. By contrast, absolute ambiguity aversion corresponds to a *weak form of compound-risk aversion*: $\delta_{\mu(Q)} \succsim Q$, that is, the mean of a compound lottery Q is preferred to Q.[10] In SORDU, the latter property requires Ψ be sufficiently small relative to ϕ. When ϕ is concave, we obtain a sharp characterization: Absolute ambiguity aversion becomes equivalent to the pessimism condition $\Psi(\alpha) \leq \alpha$ for $\alpha \in (0, 1)$.

These observations easily follow from well-known relations between risk attitudes and the parameters of the classical RDU model examined by Chew, Karni and Safra (1987), and Chateauneuf and Cohen (1994).

A widely studied example due to Machina (2009) illustrates the importance of

[9] Specifically, $V_{Sr}(P_h)$ equals the expectation of $\phi(E_{p_h}(u))$ with respect to the probability measure that assigns $\Psi(\sum_{i=j}^{m} P(p^i)) - \Psi(\sum_{i=j+1}^{m} P(p^i))$ to p^j, where $p^1, ..., p^m$ are the elements of Π^* ordered so that $E_{p_h^m}(u) \geq \cdots \geq E_{p_h^1}(u)$.

[10] Dillenberger (2010) calls this property "preference for one-shot resolution of uncertainty."

separating absolute ambiguity aversion from strong compound-risk aversion.

Reflection Example. Let $S = \{s^1, s^2, s^3, s^4\}$. The DM knows that the probabilities of the events $\{s^1, s^2\}$ and $\{s^3, s^4\}$ equal 1/2. While the probabilities of s^2 and s^3 are ambiguous, the DM also knows that they belong to the set $M := \{k/100 : k = 0, 1, ..., 50\}$. Thus,

$$\Pi^*_{ref} := \left\{p \in \Delta(S) : p(\{s^1, s^2\}) = p(\{s^3, s^4\}) = 1/2, \left(p(s^2), p(s^3)\right) \in M \times M\right\}.$$

Throughout the example, $\succsim$ denotes the DM's preference relation on $\mathbb{R}^S \cup \Delta^2(\mathbb{R})$. For any $P \in \Delta(\Pi^*_{ref})$, I write $P(k,t)$ in place of $P(\{p \in \Pi^*_{ref} : (p(s^2), p(s^3)) = (k/100, t/100)\})$, and say that P is *symmetric* if $P(k,t) = P(t,k)$ for $k, t = 0, 1, ..., 50$.

Pick two prizes x, y with $x > y > 0$, and consider the four acts defined in Table 1 below.[11]

Table 1–The Reflection Example

Act	s^1	s^2	s^3	s^4
h^5	y	x	y	0
h^6	y	y	x	0
h^7	0	x	y	y
h^8	0	y	x	y

Note that h^5 and h^8 are symmetric reflections of each other, and the same holds for h^6 and h^7. By contrast, h^5 and h^6 are significantly different; the former is more exposed to ambiguity. Indeed, in an experimental study, L'Haridon and Placido (2010) found that among ambiguity averse subjects, the most frequently observed pattern is $h^5 \prec h^6$ and $h^7 \succ h^8$. Baillon, L'Haridon and Placido (2011) showed that assuming ambiguity aversion and a symmetric second-order belief P, the smooth ambiguity model is unable to accommodate this pattern because the (expected) utility distribution induced by h^6 is a mean-preserving spread of that induced by h^5. The following claim reformulates this observation in terms of compound-risk aversion.

Claim 1. *Suppose that $\succsim$ is a SOPS preference relation represented by a symmetric second-order belief $P \in \Delta(\Pi^*_{ref})$. If $\succsim$ satisfies second-stage vN-M independence, and*

[11] Using these acts, Machina (2009) questioned the descriptive power of Schmeidler's (1989) Choquet expected utility model. In Machina's original formulation, we have $x = \$8000$, $y = \$4000$, while the state s^i ($i = 1, 2, 3, 4$) corresponds to the event of extracting a ball marked with the number i from an urn that contains 100 balls.

the strong compound-risk aversion property in expression (3)*, then we have* $h^5 \succsim h^6$*.*

Since SORDU separates absolute ambiguity aversion from strong compound-risk aversion, it is able to accommodate the pattern observed by L'Haridon and Placido (2010) with reasonable specifications:

Proposition 2. *Consider a SORDU model with* $\phi(x) \equiv x$*. If* Ψ *is not convex, then there exists a symmetric* $P \in \Delta(\Pi^*_{ref})$ *such that the associated preference* $\succsim$ *satisfies* $h^7 \sim h^6 \succ h^5 \sim h^8$*. If, in addition,* $\Psi(\alpha) \leq \alpha$ *for every* $\alpha \in (0,1)$*, then* $\succsim$ *is also ambiguity averse.*[12]

Ambiguity attitudes depending on the likelihood of a good outcome. In a recent commentary, Ellsberg (2011) notes that ambiguity aversion observed in experiments may hinge upon the likelihood of a good return. Kocher, Lahno and Trautmann (2018) test this hypothesis with a large number of subjects, and find that ambiguity seeking is indeed common when the likelihood of a good outcome is small. By contrast, ambiguity aversion prevails in classical Ellsberg-type experiments that yield a good outcome with a moderately large likelihood. Specifically, the experiments of Kocher et al. involve two urns, Urn 1 and 2, filled with colored chips of an unknown composition. Urn 1 contains up to ten different colors, while Urn 2 contains at most two different colors. More than half of the subjects prefer a bet on the color of a randomly drawn chip from Urn 1 to the corresponding risky bet with 10% chance of winning, which suggests ambiguity seeking. The opposite, ambiguity averse behavior is observed in the two-colors case: More than half of the subjects prefer a risky bet with 50% chance of winning to a bet on the color of a randomly drawn chip from Urn 2.

SORDU with an inverse S-shaped Ψ function provides a natural framework to model such behavior. This version of the model admits a number $\bar{\alpha} \in (0,1)$ such that $\Psi(\alpha) > \alpha$ for $\alpha \in (0, \bar{\alpha})$ and $\Psi(\alpha) < \alpha$ for $\alpha \in (\bar{\alpha}, 1)$. Given a bet on a particular state $\hat{s} \in S$, consider a second-order belief $P = \sum_{i=1}^n \alpha^i \delta_{p^i}$ with $p^{i+1}(\hat{s}) > p^i(\hat{s})$ for each i. The probability of the tail event $\{p^i : i \geq j\}$ under P equals $P^{j+} := \sum_{i=j}^n \alpha^i$. For a distribution p^j with $P^{j+} < \bar{\alpha}$, we have $\Psi(P^{j+}) > P^{j+}$, meaning that the tail probability P^{j+} is evaluated in an optimistic way. By contrast, $P^{j+} > \bar{\alpha}$ implies $\Psi(P^{j+}) < P^{j+}$, a pessimistic evaluation. Thus, a simultaneous increase in $P^{1+}, ..., P^{n+}$, that is, an

[12] Dillenberger and Segal (2015) show that Segal's (1987) model is compatible with all examples provided by Machina (2009, 2014). A thorough examination of SORDU in this broader context is left for future work.

upward first-order stochastic shift in winning probabilities, makes each P^{j+} more likely to fall into the pessimistic region $(\bar{\alpha}, 0]$. Consequently, a decrease in the number of states—"colors," in a typical urn experiment—may act as a catalyst for ambiguity aversion by increasing the DM's assessment of winning probabilities for the bet on the particular state $\hat{s}$.

Toward a concrete example, consider a SORDU representation with the following four properties:

i. $\phi(v) = v$ for every $v \in u(X)$.

ii. $\Psi(\alpha) - \alpha$ restricted to $(0, \bar{\alpha})$ is a positive function with a unique peak at $\bar{\alpha}/2$. (Peak 1)

iii. $-(\Psi(\alpha) - \alpha)$ restricted to $(\bar{\alpha}, 1)$ is a positive function with a unique peak at $(1 + \bar{\alpha})/2$. (Peak 2)

iv. $\bar{\alpha} < 1/2$, and Peak 2 is as large as Peak 1 in the following sense: $|\alpha - \bar{\alpha}/2| > |\alpha' - (1 + \bar{\alpha})/2|$ implies $\Psi(\alpha) - \alpha < -(\Psi(\alpha') - \alpha')$ for every $\alpha \in [0, \bar{\alpha}]$ and $\alpha' \in [\bar{\alpha}, 1]$.

Here, property (iv) serves to tilt the balance in favor of ambiguity aversion in a two-color experiment with a 50-50 lottery as the benchmark.

In the following claim, each state $s \in S$ represents a particular color we may observe on an object extracted from an urn. The bet $\hat{h}$ pays a prize x if a certain color $\hat{s}$ is observed, and a smaller prize y otherwise.

Proposition 3 . *Consider a second-order belief* $P \in \Delta^2(S)$ *such that for any* $p \in \operatorname{supp}(P)$*, the probability of extracting the color* $\hat{s}$*,* $p(\hat{s})$*, equals one of three numbers,* $\beta^1, \beta^2, \beta^3$*. Suppose* $\beta^2 \in (0, 1)$ *is the expected value of* $p(\hat{s})$*, that is,* $\beta^2 = \sum_{i=1}^{3} \beta^i P^i$*, where* $P^i := P\{p \in \Delta(S) : p(\hat{s}) = \beta^i\}$*. Assume further that* $P^2 < \bar{\alpha}$*,* $\beta^1 = 0$*, and* $\beta^3 = 1$*. Then a SORDU representation with the properties* (*i*)-(*iv*) *above implies the following:*

(a) $\hat{h} \succ \beta_2\delta_x + (1 - \beta_2)\delta_y$ *for all sufficiently small* β_2*.*

(b) $\hat{h} \prec \beta_2\delta_x + (1 - \beta_2)\delta_y$ *whenever* $\beta_2 \geq 1/2$*.*

This claim provides sufficient conditions to accommodate the findings of Kocher et al. (2018). We have ambiguity seeking when β^2—the expected probability of the winning color $\hat{s}$—is small (statement (a)). The opposite prediction obtains when β^2 is moderately large (statement (b)).

The condition $P^2 < \bar{\alpha}$ means that there is a significant amount of ambiguity concerning the true probability of $\hat{s}$. When β^2 is small, so is P^3, and consequently, both tail probabilities P^3 and $P^2 + P^3$ fall into the optimistic region $(0, \bar{\alpha})$. This immediately yields the ambiguity seeking behavior in statement (a). As for statement (b), note that holding P^2 fixed, an increase in β^2 is accompanied by an increase in P^3, making $P^2 + P^3$ more likely to fall into the pessimistic region $(\bar{\alpha}, 1)$. The weight of this tail probability in $V_{Sr}(P_h)$ is proportional to $\beta^2 - \beta^1$, while the weight of the other tail probability, P^3, is proportional to $\beta^3 - \beta^2$. Here, the particular specification with $\beta^1 = 0$ and $\beta^3 = 1$ ensures that $\beta^2 - \beta^1 \geq \beta^3 - \beta^2$ whenever $\beta^2 \geq 1/2$, so that the tail probability in the pessimistic region has a greater weight.[13]

In Proposition 3 , when $\beta^2 > 1/2$, the compound lottery $P_{\hat{h}}$ associated with P and $\hat{h}$ is negatively skewed in the sense of Dillenberger and Segal (2015). Similarly, the case $\beta^2 < 1/2$ corresponds to positively skewed noise. Assuming a quasi-concave function V on compound lotteries, Dillenberger and Segal derive aversion to negatively skewed noise as a consequence of aversion to symmetric noise. Proposition 3 above is beyond the scope of their analysis because SORDU with an inverse S-shaped function Ψ is neither quasi-concave nor quasi-convex.[14] It should also be noted that their derivation of preference for positively skewed noise relies on a not-so-straightforward differential condition on preferences, while Proposition 3 depicts a clear-cut dichotomy assuming a well-known form of the function Ψ.

In passing, let us note that the smooth ambiguity model can also generate instances of ambiguity seeking and aversion simultaneously, provided that the function ϕ is neither concave nor convex. However, this approach cannot distinguish between the effects of winning probabilities and prizes, making it impossible to prove an analogue of Proposition 3 .

5. Comparative Bayesian Reactions

Models of ambiguity aversion typically involve a negative association between the values of acts and the level of ambiguity. On a separate note, the uniqueness part of Theorem 1 tells us that distinct beliefs tend to produce distinct reactions to new information, irrespective of ambiguity attitudes. The following comparative statics exercise connects these two observations.

Proposition 4. *Let (V, P^*) and $(V, \hat{P}^*)$ be second-order Bayesian representations for*

[13] We could as well assume that $\beta^1 = \beta^2/2$ and $\beta^3 = (1 + \beta^2)/2$.

[14] In fact, SORDU is quasi-concave only if Ψ is concave (Wakker 1994, Theorem 24). Typically, this is incompatible with any form of noise aversion.

the collections $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ *and* $\{\hat{\succsim}_\Pi : \Pi \in 2^{\Pi^*}\}$, *respectively. Suppose that there exists an* $\alpha \in (0,1]$ *such that*

$$\hat{P}^* = \alpha\delta_{\mu(P^*)} + (1-\alpha)P^*. \tag{4}$$

Consider an act $h \in \mathcal{H}$ *and a set* $\Pi \in 2^{\Pi^*}$ *such that* **(a)** $\mu(P^*)_h \succ \mu(P^*|\Pi)_h$, **(b)** $\mu(P^*) \in \Pi$ *and* $P^*(\Pi \backslash \{\mu(P^*)\}) > 0$. *Assume further that either of the following two statements hold:* **(c)** $\succsim$ *is convex on* $\Delta^2(X)$ *and* $\mu(P^*|\Pi)_h \succsim_\Pi h$. **(d)** $\succsim$ *satisfies strong compound-risk aversion. Then* $(\hat{P}^*|\Pi)_h \succ (P^*|\Pi)_h$, *that is,* $h \hat{\sim}_\Pi \delta_x$ *implies* $\delta_x \succ_\Pi h$.

Equation (4) is a mean-preserving spread operation, which entails that P^* embodies a greater level of ambiguity than $\hat{P}^*$. It says that P^* can be obtained from $\hat{P}^*$ by transferring a mass from the common mean of these distributions, $\mu(P^*)$, to some other first-order distributions.[15] Proposition 4 examines when a greater level of ex-ante ambiguity, in this particular sense, deteriorates the value of an act h conditional on an event Π. If statement (c) or (d) holds in addition to statements (a) and (b), then the certainty equivalent of h according to $\hat{\succsim}_\Pi$ becomes strictly more desirable than h according to $\succsim_\Pi$.

Statement (a) means that the value of h conditional on the ex-ante mean $\mu(P^*)$ is greater than its value conditional on the interim mean $\mu(P^*|\Pi)$. Thus, the event Π is "bad news" for the act h, on average. Statement (b) ensures that the conditional beliefs $P^*|\Pi$ and $\hat{P}^*|\Pi$ are distinct. More importantly, it follows that these conditional beliefs preserve the spread relation between the ex-ante beliefs, P^* and $\hat{P}^*$. Statement (c) covers a popular version of Segal's (1987) theory that utilizes the cautious expected utility model to evaluate simple lotteries. In particular, the condition $\mu(P^*|\Pi)_h \succsim_\Pi h$ holds if the conditional preference $\succsim_\Pi$ is absolutely ambiguity averse. Finally, statement (d) captures SORDU with a concave first-stage utility index ϕ and a convex probability distortion function Ψ. Note also that Anscombe and Aumann's (1963) expected utility model satisfies both conditions (c) and (d).

Overall, Proposition 4 highlights a particular form in which the uniqueness part of Theorem 1 manifests itself: Assuming ambiguity aversion, second-order beliefs with larger spread tend to produce stronger reactions to bad news concerning the true distribution of the states.

[15] Also note that if $\alpha < 1$, then P^* and $\hat{P}^*$ attach the same relative likelihood to all pairs of first-order distributions, except $\mu(P^*)$. Evren (2019) introduced this mean-preserving spread operation in relation to Segal's (1987) model.

6. Concluding Remarks

In this paper, I modeled preferences under ambiguity with a second-order belief that is updated in a Bayesian fashion in response to new information about the true distribution of the states. Bayesian updating implies strong connections between the prior second-order belief and the DM's reactions to new information. This makes it possible to identify the DM's belief uniquely based on conditional preferences, even if the ex-ante preference relation is representable by multiple second-order beliefs.

Also, the model features a general non-expected utility function on compound lotteries. I illustrated the benefits of this generality by means of a specification that uses a rank-dependent aggregator in place of the first-stage expectation operator of the smooth ambiguity model. This and other particular specifications are easily characterized by additional axioms on preferences on compound lotteries, which provides a testing ground to compare and contrast second-order representations.

As a limitation, it should be noted that the connections between acts and compound lotteries embodied in the notion of SOPS rule out certain forms of ambiguity attitudes. Suppose, for example, that there are two possible distributions of the states, p^1 and p^2. Consider two acts h and g such that $p^1_h \succ p^2_h$ and $p^2_g \succ p^1_g$. If the DM tends to prefer compound lotteries to acts, we may have $\frac{1}{2}\delta_{p^1_h} + \frac{1}{2}\delta_{p^2_h} \succ h$ and $\frac{1}{2}\delta_{p^1_g} + \frac{1}{2}\delta_{p^2_g} \succ g$. If we take $\succsim$ as the ex-ante preference associated with the set $\Pi^* = \{p^1, p^2\}$, Consistent Expansion and Singleton Sophistication jointly rule out this mode of behavior. More generally, a second-order belief supported on $\{p^1, p^2\}$ cannot generate the pattern in question, leading to a second-order extension of the Ellsberg paradox.[16] It thus appears that second-order representations are unable to accommodate some interesting modes of behavior that involve significantly different attitudes toward ambiguity and compound risk. A thorough analysis of such behavior is beyond the scope of the present paper.

Interim signals that I have used throughout the paper can be constructed in laboratory experiments, but they have two special features that make them rare in real-life. First, the signals do not bear any information that makes necessary to update the first-order distributions. Second, the DM does not learn anything that may influence the relative likelihood of first-order distributions that are not eliminated; she only learns that a set of distributions do not contain the true distribution. I leave it to future research to explore how alternative signal structures can be incorporated into the theory.

[16] A more sophisticated example of similar nature can also be found in Halevy and Ozdenoren (2022).

The most widely studied second-order representation is the smooth ambiguity model. Seo's (2009) characterization of it focuses on a setup with compound lotteries, but does not necessitate a Bayesian updating of the second-order belief. More recently, Denti and Pomatto (2022) have provided a further characterization of this model without utilizing preferences on compound lotteries.[17] Analogous extensions of other representations that I have reported would be useful if experimental findings reveal systematic discrepancies between the valuations of acts and compound lotteries.

Appendix A.1. Proof of Theorem 1

Suppose $\{\succsim_\Pi : \Pi \in 2^{\Pi^*}\}$ is a non-constant collection that satisfies the axioms (A1)-(A5). For the present, let us also assume that each $\Pi \in 2^{\Pi^*}$ is non-null. The general case with null sets is handled at the end of the proof.

By STD(i), there exists a continuous function $V : \Delta^2(X) \mapsto \mathbb{R}$ such that $Q \succsim Q'$ iff $V(Q) \geq V(Q')$ for any $Q, Q' \in \Delta^2(X)$. Parts (iii) and (iv) of STD ensure that V is twice fsd-increasing.

Toward the proof of the "only if" part, we need to find a $P^* \in \Delta(\Pi^*)$ such that $h \sim_\Pi (P^*|\Pi)_h$ for every $h \in \mathcal{H}$ and $\Pi \in 2^{\Pi^*}$. By SiSo, without loss of generality we can assume that Π^* contains at least two distributions.

Claim A1. *For each $h \in \mathcal{H}$ and $\Pi \in 2^{\Pi^*}$, there exist $p', p'' \in \Pi$ such that $p'_h \succsim_\Pi h \succsim_\Pi p''_h$.*

Proof. Fix an $h \in \mathcal{H}$. By SiSo, the conclusion of the claim holds trivially when Π is a singleton. Inductively, fix a natural number n, and suppose that the conclusion of the claim holds for any $\Pi \in 2^{\Pi^*}$ with n elements.

Pick a $\Pi' \in 2^{\Pi^*}$ with $n+1$ elements. Let $p', p'' \in \Pi'$ be such that $p'_h \succsim p_h \succsim p''_h$ for every $p \in \Pi'$. Set $\Pi := \Pi' \backslash \{p'\}$. By the induction hypothesis, there exist $p, \hat{p} \in \Pi$ such that $p_h \succsim_\Pi h \succsim_\Pi \hat{p}_h$, or equivalently, $\delta_{p_h} \succsim_\Pi h \succsim_\Pi \delta_{\hat{p}_h}$. The sets $\{\gamma \in [0,1] : \gamma\delta_{p_h} + (1-\gamma)\delta_{\hat{p}_h} \succ_\Pi h\}$ and $\{\gamma \in [0,1] : h \succ_\Pi \gamma\delta_{p_h} + (1-\gamma)\delta_{\hat{p}_h}\}$ are relatively open in $[0,1]$ by the continuity property in STD(i). Since the interval $[0,1]$ cannot be expressed as a union of two disjoint nonempty open subsets, it follows that there exists

[17] The support of the second-order belief in Denti and Pomatto's (2022) representation satisfies a technical condition related to statistical identifiability of the true first-order distribution. Klibanoff et al. (2022) study a special case where states are infinite sequences with independently and identically distributed components. Earlier contributions in this line of research include Cerreia-Vioglio et al. (2013), and Al-Najjar and De Castro (2014).

a $\gamma \in [0,1]$ such that $h \sim_\Pi \gamma\delta_{p_h} + (1-\gamma)\delta_{\hat{p}_h}$. With $Q := \gamma\delta_{p_h} + (1-\gamma)\delta_{\hat{p}_h}$, from ME and the definition of p' it follows that $h \succsim_{\Pi'} Q$. Moreover, we have $Q \succsim_\Pi \hat{p}_h$ because $Q \sim_\Pi h \succsim_\Pi \hat{p}_h$, while $Q \succsim_\Pi \hat{p}_h$ implies $Q \succsim_{\Pi'} \hat{p}_h$ by invariance of risk preferences, STD(ii). Similarly, the definition of p'' implies $\hat{p}_h \succsim_{\Pi'} p''_h$. To summarize, we have $h \succsim_{\Pi'} Q \succsim_{\Pi'} \hat{p}_h \succsim_{\Pi'} p''_h$, and hence, $h \succsim_{\Pi'} p''_h$. From a symmetric argument it also follows that $p'_h \succsim_{\Pi'} h$. $\square$

Claim A2. (i) *For any $p, p' \in \Pi^*$ with $p \neq p'$, there exists an act $g \in \mathcal{H}$ such that $p'_g \succ p_g$.*
(ii) *There exists an act $g^* \in \mathcal{H}$ such that for any $p, p' \in \Pi^*$ with $p \neq p'$, either $p'_{g^*} \succ p_{g^*}$ or $p'_{g^*} \prec p_{g^*}$.*

Proof. For any $p, p' \in \Pi^*$ with $p \neq p'$, there exists an $s' \in S$ such that $p'(s') > p(s')$. Pick some $x, y \in X$ with $x > y$. Define an act g as $g(s') := x$ and $g(s) := y$ for $s \in S \backslash \{s'\}$. Then $p'_g = p'(s')\delta_x + (1-p'(s'))\delta_y >_{fsd} p(s')\delta_x + (1-p(s'))\delta_y = p_g$, which implies $p'_g \succ p_g$ by STD(iii). This proves statement (i).

To prove (ii), suppose S has m elements, $s^1, s^2, ..., s^m$. Select a pair of numbers x, y in the interior of X with $x > y$. Define an act g^1 as $g(s^1) := x$ and $g(s) := y$ for $s \in S \backslash \{s^1\}$. Then, as in the first part of the proof, $p_{g^1} \sim p'_{g^1}$ implies $p(s^1) = p'(s^1)$ for every $p, p' \in \Pi^*$.

I claim that for any integer k with $1 \leq k < m$, there exists an act g^k, which maps S into the interior of X, such that: $g^k(s^i) = g^k(s^{k+1})$ for $i = k+1, ..., m$, and for every $p, p' \in \Pi^*$

$$p_{g^k} \sim p'_{g^k} \quad \Rightarrow \quad p(s^i) = p'(s^i) \ \forall i \in \{1, ..., k\}.$$

We have already established this claim for $k = 1$. So, let $m \geq 3$, and suppose that there exists a g^{k_o} that satisfies the desired properties for some k_o with $1 \leq k_o < m - 1$.

Set $k := k_o + 1$ and $z^k := g^{k_o}(s^k)$. Given an $\varepsilon > 0$, select a pair of numbers, $z, \hat{z}$, in the interior of X such that $z^k - \varepsilon < \hat{z} < z < z^k$. Define an act g^k as $g^k(s^i) := g^{k_o}(s^i)$ for $i \leq k_o$, $g^k(s^k) := z$, and $g^k(s^i) := \hat{z}$ for $i > k$. Then $\lim_{\varepsilon \to 0} p_{g^k} = p_{g^{k_o}}$ for any $p \in \Pi^*$ because $\lim_{\varepsilon \to 0} g^k(s) = g^{k_o}(s)$ for every $s \in S$. Since Π^* is finite and $\succsim$ is continuous, it follows that there exists a sufficiently small $\varepsilon > 0$ such that $p_{g^{k_o}} \succ p'_{g^{k_o}}$ implies $p_{g^k} \succ p'_{g^k}$ for any $p, p' \in \Pi^*$. Put differently, whenever $p_{g^k} \sim p'_{g^k}$, we also have $p_{g^{k_o}} \sim p'_{g^{k_o}}$.

Suppose $p_{g^k} \sim p'_{g^k}$ for some $p, p' \in \Pi^*$. Then $p_{g^{k_o}} \sim p'_{g^{k_o}}$, which implies $p(s^i) = p'(s^i)$ for $i = 1, ..., k_o$ by our choice of g^{k_o}. Hence, $p'_{g^k} = \sum_{i \leq k_o} p(s^i)\delta_{g^k(s^i)} + p'(s^k)\delta_z +$

$\left(1 - p'(s^k) - \sum_{i \leq k_o} p(s^i)\right) \delta_{\hat{z}}$. Since $z > \hat{z}$, upon expanding p_{g^k} in a similar fashion, we see that $p(s^k) > p'(s^k)$ implies $p_{g^k} >_{fsd} p'_{g^k}$, which contradicts the assumption that $p_{g^k} \sim p'_{g^k}$. Similarly, we cannot have $p'(s^k) > p(s^k)$. Thus, $p(s^i) = p'(s^i)$ for $i = 1, ..., k$, as we seek.

Inductively, we can find an act $g^* := g^{m-1}$ such that $p_{g^*} \sim p'_{g^*}$ implies $p(s^i) = p'(s^i)$ for $i = 1, ..., m-1$, that is, $p = p'$. $\square$

Fix a set $\Pi := \{p, p'\}$ for some distinct $p, p' \in \Pi^*$. Pick an act $g \in \mathcal{H}$ with $p'_g \succ p_g$. Claim A1 implies $p'_g \succsim_\Pi g \succsim_\Pi p_g$. The proof of the claim also shows that there is a $\gamma(p', p) \in [0, 1]$ such that $g \sim_\Pi \gamma(p', p)\delta_{p'_g} + (1 - \gamma(p', p))\, \delta_{p_g}$. In fact, $\gamma(p', p)$ is the unique number $\gamma \in [0, 1]$ that satisfies $g \sim_\Pi \gamma\delta_{p'_g} + (1 - \gamma)\, \delta_{p_g}$ because $V(\gamma\delta_{p'_g} + (1 - \gamma)\delta_{p_g})$ is increasing in γ. Moreover, $\gamma(p', p) < 1$, for otherwise we would have $g \sim_\Pi p'_g$ while $p'_g \succ p_g$, contradicting the assumption that p is non-null.

Let us now show that $h \sim_\Pi \gamma(p', p)\delta_{p'_h} + (1 - \gamma(p', p))\, \delta_{p_h}$ for every $h \in \mathcal{H}$. By part (1) of Definition 3, the degenerate distribution δ_p is pre-consistent on $\{p\}$. Consequently, part (2) of the definition implies that the distribution $P := \gamma(p', p)\delta_{p'} + (1 - \gamma(p', p))\, \delta_p$ is also pre-consistent on Π. Since $h \sim_{\{p'\}} \delta_{p'_h}$ for any $h \in \mathcal{H}$, invoking CE with $p^1 = p^2 = p'$ yields $h \sim_\Pi P(p')\delta_{p'_h} + P(p)\delta_{p_h} = \gamma(p', p)\delta_{p'_h} + (1 - \gamma(p', p))\delta_{p_h}$, as we seek.

Symmetrically, we also have $h \sim_\Pi \gamma(p, p')\delta_{p_h} + (1 - \gamma(p, p'))\delta_{p'_h}$ for every $h \in \mathcal{H}$. From the aforementioned uniqueness property of the function $\gamma(\cdot, \cdot)$, it follows that $\gamma(p, p') = 1 - \gamma(p', p)$. As $\gamma(p, p')$ and $\gamma(p', p)$ are both less than 1, we get $0 < \gamma(p', p) < 1$.

Thereby, we have shown that for any $\Pi \in 2^{\Pi^*}$ with two elements, there exists a pre-consistent distribution $P \in \Delta(\Pi)$ such that (a) $P(p) > 0$ for every $p \in \Pi$; and (b) $h \sim_{\Pi'} (P|\Pi')_h$ for every $h \in \mathcal{H}$ and $\Pi' \in 2^\Pi$. Inductively, fix an integer $n \geq 2$, and suppose that for every $\Pi \in 2^{\Pi^*}$ with n elements, there exists a pre-consistent distribution $P \in \Delta(\Pi)$ that satisfies the properties (a) and (b).

Consider a set $\Pi \in 2^{\Pi^*}$ with $n + 1$ elements. Given an act g^* as in Claim A2(ii), let p' denote the element of Π such that $p'_{g^*} \succ p_{g^*}$ for every $p \in \hat{\Pi} := \Pi \backslash \{p'\}$. By the induction hypothesis, there exists a pre-consistent distribution $\hat{P} \in \Delta(\hat{\Pi})$ such that $\hat{P}(p) > 0$ for every $p \in \hat{\Pi}$, and $h \sim_{\Pi'} (\hat{P}|\Pi')_h$ for every $h \in \mathcal{H}$ and $\Pi' \in 2^{\hat{\Pi}}$. In particular, $g^* \sim_{\hat{\Pi}} \hat{P}_{g^*}$, and hence, ME implies $g^* \succsim_\Pi \hat{P}_{g^*}$. Moreover, we have $p'_{g^*} \succsim_\Pi g^*$ by Claim A1. Thus, as in the proof of Claim A1, the continuity property in STD(i) implies that $g^* \sim_\Pi \gamma^*\delta_{p'_{g^*}} + (1 - \gamma^*)\hat{P}_{g^*}$ for some $\gamma^* \in [0, 1]$. Since $\hat{P}$ is pre-consistent on $\hat{\Pi}$, it follows that so is the distribution $P := \gamma^*\delta_{p'} + (1 - \gamma^*)\hat{P}$ on Π. To complete the induction, we need to prove that P satisfies the properties (a) and (b).

Fix a $p \in \Pi \backslash \{p'\}$. Let $\gamma' := \gamma(p', p)$ so that $h \sim_{\{p,p'\}} \gamma' \delta_{p'_h} + (1-\gamma')\, \delta_{p_h}$ for every $h \in \mathcal{H}$. Since P is pre-consistent on Π, from CE it follows that

$$h \sim_{\Pi} P(\{p, p'\}) \left(\gamma' \delta_{p'_h} + (1-\gamma')\, \delta_{p_h}\right) + \textstyle\sum_{\hat{p} \in \Pi \backslash \{p,p'\}} P(\hat{p}) \delta_{\hat{p}_h} \quad \forall h \in \mathcal{H}. \tag{5}$$

Let $Q^* := P(\{p,p'\})(\gamma' \delta_{p'_{g^*}} + (1-\gamma')\, \delta_{p_{g^*}}) + \sum_{\hat{p} \in \Pi \backslash \{p,p'\}} P(\hat{p}) \delta_{\hat{p}_{g^*}}$. By (5), we have $g^* \sim_{\Pi} Q^*$. Moreover, $g^* \sim_{\Pi} \gamma^* \delta_{p'_{g^*}} + (1-\gamma^*) \hat{P}_{g^*} = P_{g^*}$ by construction, and hence, $Q^* \sim P_{g^*}$. Note also that $Q^*(q) = P_{g^*}(q)$ for each $q \in \Delta(X) \backslash \{p'_{g^*}, p_{g^*}\}$. As $p'_{g^*} \succ p_{g^*}$, from part (iv) of STD it follows that $Q^* = P_{g^*}$, or equivalently,

$$P = P(\{p, p'\}) \left(\gamma' \delta_{p'} + (1-\gamma')\, \delta_{p}\right) + \textstyle\sum_{\hat{p} \in \Pi \backslash \{p,p'\}} P(\hat{p}) \delta_{\hat{p}}. \tag{6}$$

This equation implies that $P(p)$ and $P(p')$ are both positive numbers because γ', $1-\gamma'$, and $P(\{p,p'\}) = \gamma^* + (1-\gamma^*) \hat{P}(p)$ are all positive. Hence, γ^* belongs to $(0,1)$, and $P(\hat{p}) > 0$ for every $\hat{p} \in \Pi$, which verifies the condition (a).

By equations (5) and (6), we have $h \sim_{\Pi} P_h$ for every $h \in \mathcal{H}$. Pick an arbitrary $h \in \mathcal{H}$ and $\Pi' \in 2^{\Pi}$ with $\Pi' \neq \Pi$. The next step is to show that $h \sim_{\Pi'} (P|\Pi')_h$.

If $\Pi' \subseteq \hat{\Pi}$, we have $P|\Pi' = \hat{P}|\Pi'$ by definition of P, and the desired conclusion follows from the induction hypothesis $h \sim_{\Pi'} (\hat{P}|\Pi')_h$. So, without loss of generality we can assume that $p' \in \Pi'$.

Since Π' is a proper subset of Π, by the induction hypotheses there exists a pre-consistent distribution $P' \in \Delta(\Pi')$ such that $P'(p) > 0$ for every $p \in \Pi'$, and $\hat{h} \sim_{\Pi''} (P'|\Pi'')_{\hat{h}}$ for every $\hat{h} \in \mathcal{H}$ and $\Pi'' \in 2^{\Pi'}$. In particular, $h \sim_{\Pi'} P'_h$, and hence, it suffices to show that $P|\Pi' = P'$.

Pick any $p \in \Pi' \backslash \{p'\}$. By equation (6), $P(p)/P(p') = (1-\gamma(p',p))\,/\gamma(p',p)$. Moreover, $g^* \sim_{\{p,p'\}} (P'|\{p,p'\})_{g^*}$ by definition of P', which implies $P'|\{p,p'\} = \gamma(p',p)\delta_{p'} + (1-\gamma(p',p))\delta_p$. Thus,

$$\frac{P|\Pi'(p)}{P|\Pi'(p')} = \frac{P(p)}{P(p')} = \frac{1-\gamma(p',p)}{\gamma(p',p)} = \frac{P'|\{p,p'\}(p)}{P'|\{p,p'\}(p')} = \frac{P'(p)}{P'(p')}, \tag{7}$$

where the first and last equalities hold by definition of a Bayesian update. Since p is an arbitrary element of $\Pi'\backslash\{p'\}$, it follows that

$$\frac{1-P|\Pi'(p')}{P|\Pi'(p')} = \sum_{p\in\Pi'\backslash\{p'\}} \frac{P|\Pi'(p)}{P|\Pi'(p')} = \sum_{p\in\Pi'\backslash\{p'\}} \frac{P'(p)}{P'(p')} = \frac{1-P'(p')}{P'(p')},$$

which means $P|\Pi'(p') = P'(p')$. Then invoking the equality (7) once again yields $P|\Pi'(p) = P'(p)$ for every $p \in \Pi'\backslash\{p'\}$. Hence, $P|\Pi' = P'$, as we seek.

The inductive procedure above leads to a $P^* \in \Delta(\Pi^*)$ such that $P^*(p) > 0$ for every $p \in \Pi^*$, and $h \sim_\Pi (P^*|\Pi)_h$ for every $h \in \mathcal{H}$ and $\Pi \in 2^{\Pi^*}$. This distribution P^* coupled with the function V gives us a second-order Bayesian representation, under the assumption that each $\Pi \in 2^{\Pi^*}$ is non-null.

Suppose now that some subsets of Π^* are null. Let Π^+ denote the set of non-null distributions, $\Pi^+ := \{p \in \Pi^* : \{p\} \text{ is non-null}\}$. By definitions, a set $\Pi \in 2^{\Pi^*}$ is null iff $\Pi \subseteq \Pi^*\backslash\Pi^+$. If every nonempty subset of Π^* were null, NI(i) would imply $\succsim_\Pi = \succsim$ for every $\Pi \in 2^{\Pi^*}$. Since the collection $\{\succsim_\Pi\colon \Pi \in 2^{\Pi^*}\}$ is non-constant, it follows that Π^* has some non-null subsets, and Π^+ is non-empty.

Upon replacing Π^* with Π^+ in the first part of the proof, we obtain a $P^* \in \Delta(\Pi^*)$ such that $\text{supp}(P^*) = \Pi^+$, and $h \sim_\Pi (P^*|\Pi)_h$ for every $h \in \mathcal{H}$ and $\Pi \in 2^{\Pi^+}$. Since $\text{supp}(P^*) = \Pi^+$, we have $P^*(\Pi) > 0$ iff $\Pi \cap \Pi^+ \neq \emptyset$ for any $\Pi \in 2^{\Pi^*}$. As the condition $\Pi \cap \Pi^+ \neq \emptyset$ characterizes non-null sets, we see that $\Pi \in 2^{\Pi^*}$ is non-null iff $P^*(\Pi) > 0$. Moreover, by NI(ii), $\succsim_{\Pi\cap\Pi^+} = \succsim_\Pi$ for any non-null $\Pi \in 2^{\Pi^*}$, while $\text{supp}(P^*) = \Pi^+$ implies $P^*|\Pi = P^*|\Pi \cap \Pi^+$. It follows that $h \sim_{\Pi\cap\Pi^+} (P^*|\Pi \cap \Pi^+)_h = (P^*|\Pi)_h$ for any $h \in \mathcal{H}$ and $\Pi \in 2^{\Pi^*}$ with $P^*(\Pi) > 0$. Since $\succsim_{\Pi\cap\Pi^+} = \succsim_\Pi$, it also follows that $h \sim_\Pi (P^*|\Pi)_h$, as demanded by a second-order Bayesian representation. This proves the "only if" part of the theorem.

In what follows, (V, P^*) denotes a second-order Bayesian representation for the collection $\{\succsim_\Pi\colon \Pi \in 2^{\Pi^*}\}$.

To establish the uniqueness of P^*, suppose $\{\succsim_\Pi\colon \Pi \in 2^{\Pi^*}\}$ admits another second-order Bayesian representation $(\hat{V}, P)$ with $P \neq P^*$. Then there exist some $p, p' \in \Pi^*$ such that $P^*(p) > P(p)$ and $P^*(p') < P(p')$. Pick a $g \in \mathcal{H}$ with $p'_g \succ p_g$. Note that $(P^*|\{p,p'\})_g \prec (P|\{p,p'\})_g$ because we have $P^*|\{p,p'\}(p') < P|\{p,p'\}(p')$, and $V(\alpha\delta_{p'_g} + (1-\alpha)\delta_{p_g})$ is increasing in $\alpha \in [0,1]$. Moreover, $P^*(\{p,p'\}) \geq P^*(p) > 0$ and $P(\{p,p'\}) \geq P(p') > 0$. Hence, the representations require $g \sim_{\{p,p'\}} (P^*|\{p,p'\})_g$ and $g \sim_{\{p,p'\}} (P|\{p,p'\})_g$, which contradict the condition $(P^*|\{p,p'\})_g \prec (P|\{p,p'\})_g$.

The next step is to characterize the null sets. Pick any $p \in \Pi^*$. If $P^*(p) = 0$, there exists a $p' \in \Pi^*\backslash\{p\}$ with $P^*(p') > 0$. Then $P^*|\{p, p'\} = \delta_{p'}$, while $p'_h \succ p_h$ for some $h \in \mathcal{H}$. Moreover, $h \sim_{\{p,p'\}} p'_h$ by the representation. Hence, $\{p\}$ is null whenever $P^*(p) = 0$. Conversely, suppose $\{p\}$ is null so that $p'_h \succ p_h$ and $h \sim_{\{p,p'\}} p'_h$ for some $p' \in \Pi^*$ and $h \in \mathcal{H}$. Assume by contradiction that $P^*(p) > 0$. With $\alpha' := (P^*|\{p, p'\})(p')$, the representation implies $h \sim_{\{p,p'\}} \alpha'\delta_{p'_h} + (1-\alpha')\delta_{p_h}$. Since h is indifferent to both $\alpha'\delta_{p'_h} + (1-\alpha')\delta_{p_h}$ and $\delta_{p'_h}$ conditional on $\{p, p'\}$, we must then have $V(\alpha'\delta_{p'_h} + (1-\alpha')\delta_{p_h}) = V(\delta_{p'_h})$. This contradicts the assumption that V is twice fsd-increasing because $P^*(p) > 0$ implies $\alpha' < 1$. So, we have shown that $\{p\}$ is null iff $P^*(p) = 0$. From the definition of a null set, it also follows that a set $\Pi \in 2^{\Pi^*}$ is null iff $P^*(\Pi) = 0$.

Given this characterization of null sets, it is a routine exercise to verify the necessity of the axioms (A1)-(A3) and (A5). To prove necessity of (A4), first we need to show that $P = P^*|\Pi$ for any pre-consistent distribution $P \in \Delta(\Pi)$ on a (non-null) set $\Pi \in 2^{\Pi^*}$. This holds trivially when Π is a singleton. Given an $n \in \mathbb{N}$, suppose that the claim holds for any pre-consistent distribution on any set $\hat{\Pi}$ with n elements. Pick a set $\Pi \in 2^{\Pi^*}$ with $n+1$ elements and a pre-consistent distribution $P \in \Delta(\Pi)$. By Definition 3, there exist a $p' \in \Pi$, an act g with $p'_g \succ p_g$ for every $p \in \Pi\backslash\{p'\}$, a distribution $\hat{P} \in \Delta(\Pi\backslash\{p'\})$ that is pre-consistent on $\Pi\backslash\{p'\}$, and a number $\gamma \in [0, 1]$ such that

$$g \sim_\Pi \gamma\delta_{p'_g} + (1-\gamma)\hat{P}_g \quad \text{and} \quad P = \gamma\delta_{p'} + (1-\gamma)\hat{P}. \tag{8}$$

With $\hat{\Pi} := \Pi\backslash\{p'\}$, the induction hypothesis implies $\hat{P} = P^*|\hat{\Pi}$. Moreover, by definition of a Bayesian update, $P^*|\Pi = \alpha'\delta_{p'} + (1-\alpha')P^*|\hat{\Pi}$ where $\alpha' := P^*|\Pi(p')$. As $\hat{P} = P^*|\hat{\Pi}$, it follows that $P^*|\Pi = \alpha'\delta_{p'} + (1-\alpha')\hat{P}$. Thus, $g \sim_\Pi (P^*|\Pi)_g = \alpha'\delta_{p'_g} + (1-\alpha')\hat{P}_g$ by the representation, while the left hand side of (8) implies $\alpha'\delta_{p'_g} + (1-\alpha')\hat{P}_g \sim \gamma\delta_{p'_g} + (1-\gamma)\hat{P}_g$. Since $p'_g \succ p_g$ for each $p \in \Pi\backslash\{p'\}$, a usual monotonicity argument yields $\alpha' = \gamma$. That is, $P^*|\Pi = P$.

To verify (A4), let $P \in \Delta(\Pi)$ be a pre-consistent distribution on a set Π. As we have just seen, this implies $P^*|\Pi = P$. Let $p^1, p^2 \in \Pi$ and assume $\{p^1, p^2\}$ is non-null. With $\gamma := P^*|\{p^1, p^2\}(p^1)$, the definition of a Bayesian update implies $P^*|\Pi = P^*|\Pi(\{p^1, p^2\})(\gamma\delta_{p^1} + (1-\gamma)\delta_{p^2}) + \sum_{p\in\Pi\backslash\{p^1,p^2\}} P^*|\Pi(p)\delta_p$. Fix an $h \in \mathcal{H}$, and set $Q_{\alpha,h} := \alpha\delta_{p^1_h} + (1-\alpha)\delta_{p^2_h}$ for $\alpha \in [0, 1]$. By the representation, we have $h \sim_\Pi P(\{p^1, p^2\})Q_{\gamma,h} + \sum_{p\in\Pi\backslash\{p^1,p^2\}} P(p)\delta_{p_h}$ on the one hand, and $h \sim_{\{p^1,p^2\}} Q_{\gamma,h}$ on the other. Hence, assuming $p^1_h \succ p^2_h$, the usual monotonicity property of V implies $h \succsim_{\{p^1,p^2\}} Q_{\alpha,h}$ iff $\gamma \geq \alpha$ iff $h \succsim_\Pi P(\{p^1, p^2\})Q_{\alpha,h} + \sum_{p\in\Pi\backslash\{p^1,p^2\}} P(p)\delta_{p_h}$ for any

$\alpha \in [0,1]$. The case $p_h^1 \prec p_h^2$ is analogous. Finally, $p_h^1 \sim p_h^2$ implies $h \sim_{\{p^1,p^2\}} Q_{\alpha,h}$ and $h \sim_\Pi P(\{p^1,p^2\})Q_{\alpha,h} + \sum_{p\in\Pi\setminus\{p^1,p^2\}} P(p)\delta_{p_h}$ for any $\alpha \in [0,1]$. This completes the proof of Theorem 1.

Appendix A.2. Remaining Proofs

Proof of Proposition 1. In view of Theorem 1, to prove the "if" part of statements (i)-(iii) it suffices to show that $\succsim := \succsim_{\Pi^*}$ restricted to $\Delta^2(X)$ admits a utility function V that has the desired form, given the additional assumptions on $\succsim$ specified in the corresponding statement.

The rule $\psi : Q_o \to q_o$, with $q_o(x) := Q_o(\delta_x)$ for $x \in X$, defines a one-to-one function from Δ_o onto $\Delta(X)$, which is also affine in the sense that $\psi(\alpha Q_o + (1-\alpha)Q'_o) = \alpha\psi(Q_o) + (1-\alpha)\psi(Q'_o)$ for $Q_o, Q'_o \in \Delta_o$ and $\alpha \in (0,1)$. The inverse of ψ is a continuous function that maps $q \in \Delta(X)$ to the compound lottery $\psi^{-1}q := \sum_{x\in X} q(x)\delta_{\delta_x}$.

Define a binary relation $\succsim_o$ on $\Delta(X)$ as $q \succsim_o q'$ iff $\psi^{-1}q \succsim \psi^{-1}q'$. Since the function ψ^{-1} is continuous, the relation $\succsim_o$ inherits continuity of $\succsim$. That is, the continuity property in part (i) of STD ensures that $\{q \in \Delta(X) : q \succ_o q'\}$ and $\{q \in \Delta(X) : q' \succ_o q\}$ are open subsets of $\Delta(X)$ for each $q' \in \Delta(X)$. Also, $q >_{fsd} q'$ implies $q \succ_o q'$ by parts (iii) and (iv) of STD.

To prove part (iii) of the proposition, suppose that $\succsim$ satisfies second-stage vN-M independence and first-stage weak commutativity. By the latter property, it is easily checked that $\succsim_o$ satisfies the weak commutativity axiom as stated in Chew (1989).[18] Thus, by Chew's Theorem 1, the relation $\succsim_o$ admits a utility function $V_o : \Delta(X) \to \mathbb{R}$ defined as

$$V_o(q) := v_o(x^1) + \textstyle\sum_{j=2}^n \left(v_o(x^j) - v_o(x^{j-1})\right) \Psi(\sum_{i=j}^m q(x^i)). \tag{9}$$

Here $x^n \geq \cdots \geq x^1$ are the points in the support of the lottery $q \in \Delta(X)$, while $v_o : X \to \mathbb{R}$ and $\Psi : [0,1] \to [0,1]$ are increasing and continuous functions with $\Psi(0) = 0$ and $\Psi(1) = 1$.

Let $x(q) \in X$ denote the certainty equivalent of $q \in \Delta(X)$ defined by the property $\delta_{x(q)} \sim q$ (i.e., $\delta_{\delta_{x(q)}} \sim \delta_q$); the existence of certainty equivalents is a well-known

[18] That is, if $\delta_x \sim_o q = \sum_{i=1}^n \alpha^i \delta_{x^i}$ and $\delta_{\hat{x}} \sim_o \hat{q} = \sum_{i=1}^n \alpha^i \delta_{\hat{x}^i}$ for some $\{x^1, ..., x^n, \hat{x}^1, ..., \hat{x}^n\} \subseteq X$ and $\{\alpha^1, ..., \alpha^n\} \subseteq (0,1)$ with $\sum_{i=1}^n \alpha^i = 1$, $x^n \geq \cdots \geq x^1$, $\hat{x}^n \geq \cdots \geq \hat{x}^1$, and $x^n \geq \hat{x}^n, ..., x^1 \geq \hat{x}^1$, then $\lambda \in (0,1)$ and $\delta_{\bar{x}^i} \sim_o \lambda\delta_{x^i} + (1-\lambda)\delta_{\hat{x}^i}$ for $i = 1, ..., n$ imply $\lambda\delta_x + (1-\lambda)\delta_{\hat{x}} \sim_o \sum_{i=1}^n \alpha^i \delta_{\bar{x}^i}$.

consequence of STD(i) and (iii). Define a further function $V : \Delta^2(X) \to \mathbb{R}$ as $V(Q) := V_o(\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)})$. Since V_o represents $\succsim_o$, for any $Q, Q' \in \Delta^2(X)$, we have $V(Q) \geq V(Q')$ iff $\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)} \succsim_o \sum_{q\in\Delta(X)} Q'(q)\delta_{x(q)}$. By definition of $\succsim_o$, this means

$$V(Q) \geq V(Q') \quad \Leftrightarrow \quad \textstyle\sum_{q\in\Delta(X)} Q(q)\delta_{\delta_{x(q)}} \succsim \sum_{q\in\Delta(X)} Q'(q)\delta_{\delta_{x(q)}}.$$

Note also that $\sum_{q\in\Delta(X)} Q(q)\delta_{\delta_{x(q)}} \sim \sum_{q\in\Delta(X)} Q(q)\delta_q = Q$ by part (iv) of STD, and similarly, $\sum_{q\in\Delta(X)} Q'(q)\delta_{\delta_{x(q)}} \sim Q'$. Thus, it follows that $V(Q) \geq V(Q')$ iff $Q \succsim Q'$; that is, the function V represents $\succsim$ on $\Delta^2(X)$.

For $Q = \sum_{i=1}^n Q(q^i)\delta_{q^i}$ with $q^n \succsim \cdots \succsim q^1$, the definition of V and equation (9) jointly imply

$$V(Q) = v_o(x(q^1)) + \textstyle\sum_{j=2}^n \left(v_o(x(q^j)) - v_o(x(q^{j-1}))\right) \Psi(\sum_{i=j}^m Q(q^i)). \tag{10}$$

In particular, $v(q) := V(\delta_q) = v_o(x(q))$ for each $q \in \Delta(X)$.

By second-stage vN-M independence, there exists an expected utility function $u : \Delta(X) \to \mathbb{R}$ that represents the restriction of $\succsim$ to $\Delta(X)$, while Lemma 1 delivers a continuous and increasing function $\phi : u(X) \to \mathbb{R}$ such that $v = \phi \circ u$. Hence, equation (10) can equivalently be written as

$$V(Q) = \phi(u(q^1)) + \textstyle\sum_{j=2}^n \left(\phi(u(q^j)) - \phi(u(q^{j-1}))\right) \Psi(\sum_{i=j}^m Q(q^i)).$$

This proves the "if" part of statement (iii).

The "if" part of statement (i) is proved similarly, by replacing first-stage weak commutativity with first-stage vN-M independence, and appealing to an expected utility theorem in place of Chew's theorem.

To prove the "if" part of statement (ii), suppose that $\succsim$ satisfies time neutrality, so that $Q_o \sim \delta_{q_o}$ for each $Q_o \in \Delta_o$. Pick any function $u : \Delta(X) \to \mathbb{R}$ that represents the restriction of $\succsim$ to $\Delta(X)$ so that $u(q) \geq u(q')$ iff $\delta_q \succsim \delta_{q'}$ for any $q, q' \in \Delta(X)$. Let u^{-1} denote the inverse of the restriction of u to X. Since the increasing transformation $q \to u^{-1}(u(q))$ is ordinally equivalent to u on $\Delta(X)$, without loss of generality we can assume that $u(x) := u(\delta_x)$ is equal to x for each $x \in X$. Then $u(q) = u(\delta_{x(q)}) = x(q)$ for any $q \in \Delta(X)$.

Time neutrality combined with the definitions of u and $\succsim_o$ imply $q \succsim_o q'$ iff $\sum_{x\in X} q(x)\delta_{\delta_x} \succsim \sum_{x\in X} q'(x)\delta_{\delta_x}$ iff $\delta_q \succsim \delta_{q'}$ iff $u(q) \geq u(q')$. That is, u also represents the relation $\succsim_o$ on $\Delta(X)$. Hence, if we let $V_o(q) := u(q)$, the function $V(Q) := V_o(\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)}) = u(\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)})$ represents the relation $\succsim$ on $\Delta^2(X)$, as we have seen in the first part of the proof. Finally, note that $u(\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)}) = x(\sum_{q\in\Delta(X)} Q(q)\delta_{x(q)})$ for each $Q \in \Delta^2(X)$, as we seek.

I omit the proofs of the "only if" statements. □

Proof of Claim 1. Denote by k and t generic elements of the set $\{0, 1, ..., 50\}$. Given a distribution $p \in \Pi^*_{ref}$ with $p(s^2) = k/100$ and $p(s^3) = t/100$, let us write $q^i(k,t)$ in place of p_{h^i}. Then, where $Q^i := \sum_{k,t} P(k,t)\delta_{q^i(k,t)}$, we have $h^i \sim Q^i$ for $i = 5, 6$.

Pick an expected utility function $u : \Delta(X) \to \mathbb{R}$ that represents $\succsim$ restricted to $\Delta(X)$. Let us write $u^i(k,t)$ in place of $u(q^i(k,t))$, and u_z in place of $u(z)$ for $z \in X$. Then we have $u^6(k,t) = (t(u_x - u_0) + 50(u_y + u_0))/100$, and $u^5(k,t) = (k(u_x - u_y) + t(u_y - u_0) + 50(u_y + u_0))/100$ for any k and t.

Set $\lambda := (u_y - u_0)/(u_x - u_0)$, $\tilde{q}^5(k,t) := \lambda q^6(k,t) + (1-\lambda)q^6(t,k)$ and $Q(k,t) := \lambda\delta_{q^6(k,t)} + (1-\lambda)\delta_{q^6(t,k)}$. Note that $u(\tilde{q}^5(k,t)) = u^5(k,t)$ for any k and t. Thus, STD(iv) and the symmetry of P jointly imply

$$
\begin{aligned}
Q^5 &\sim \sum_k P(k,k)\delta_{q^5(k,k)} + \sum_{k>t}\left(P(k,t)\delta_{\tilde{q}^5(k,t)} + P(k,t)\delta_{\tilde{q}^5(t,k)}\right), \quad \text{and} \\
Q^6 &= \sum_k P(k,k)\delta_{q^6(k,k)} + \sum_{k>t}\left(P(k,t)Q(k,t) + P(k,t)Q(t,k)\right).
\end{aligned}
$$

Since $u^5(k,k) = u^6(k,k)$ and $\tilde{q}^5(k,t) = \mu(Q(k,t))$ for any k and t, it follows that $Q^5 \succsim Q^6$ if $\succsim$ satisfies strong compound-risk aversion. □

Proof of Proposition 2. Since $\phi(x) \equiv x$ is a concave function, $\succsim$ is ambiguity averse iff $\Psi(\alpha) \le \alpha$ for every $\alpha \in (0,1)$ (cf. Chateauneuf and Cohen 1994, Corollary 1).

Suppose now that Ψ is not convex. As Ψ is continuous, this means that there exist $\alpha, \beta \in (0,1)$ such that $\Psi((\alpha+\beta)/2) > (\Psi(\alpha) + \Psi(\beta))/2$. Without loss of generality, let us assume $\alpha > \beta$.

Note that $u^5(50,50) \ge \max\{u^5(50,0), u^5(0,50)\}$ and $\min\{u^5(50,0), u^5(0,50)\} \ge u^5(0,0)$, using the notation in the proof of Claim 1. Moreover, $u^5(50,0) \ge u^5(0,50)$ iff $u_x + u_0 \ge 2u_y$.

Define $P \in \Delta(\Pi^*_{ref})$ as $P(0,0) = 1-\alpha$, $P(0,50) = P(50,0) = (\alpha-\beta)/2$, $P(50,50) = \beta$, and $P(k,t) = 0$ for all other $k, t \in \{0, 1, ..., 50\}$.[19] Using the aforementioned ranking

[19] If $\alpha + \beta = 1$, then P also satisfies the additional symmetry condition $P(50-k, 50-t) = P(k,t)$ for every k and t.

of $u^5(k,t)$, it is easily checked that

$$V_{Sr}(\sum_{k,t} P(k,t)\delta_{q^6(k,t)}) - V_{Sr}(\sum_{k,t} P(k,t)\delta_{q^5(k,t)}) = \gamma(\Psi(\tfrac{\alpha+\beta}{2}) - \tfrac{\Psi(\alpha)+\Psi(\beta)}{2}),$$

where $\gamma := u_y - u_0$ if $u_x + u_0 \geq 2u_y$, and $\gamma := u_x - u_y$ otherwise. This shows that $h^6 \succ h^5$.

Finally, note that $q^5(k,t) = q^8(t,k)$ and $q^6(k,t) = q^7(t,k)$ for every k and t. Thus, the symmetry of P implies $h^5 \sim h^8$ and $h^6 \sim h^7$. $\square$

Proof of Proposition 3 . Let $u_{xy} := u(x) - u(y)$ and $u^i := u(y) + \beta^i u_{xy}$ for $i = 1,2,3$. Given any $p \in \Delta(S)$ with $p(\hat{s}) = \beta^i$, we have $p_{\hat{h}} = \beta^i\delta_x + (1-\beta^i)\delta_y$, while $E_{p_{\hat{h}}}(u) = u^i$. Thus, $V_{Sr}(P_{\hat{h}}) = u^1 + (u^2 - u^1)\Psi(P^2+P^3) + (u^3 - u^2)\Psi(P^3)$. Since $v(\beta^2\delta_x + (1-\beta^2)\delta_y) = u^2$, we need to determine the sign of the number $D := V_{Sr}(P_{\hat{h}}) - u^2$.

Note that $u^i - u^j = (\beta^i - \beta^j)u_{xy}$ for any $i,j \in \{1,2,3\}$. It easily follows that $D = (\beta^2 - \beta^1)u_{xy}(\Psi(P^2+P^3) - 1) + (\beta^3 - \beta^2)u_{xy}\Psi(P^3)$. Since $u_{xy} > 0$, without loss of generality we can set $u_{xy} = 1$.

By assumption, $\beta^2 = \sum_{i=1}^3 \beta^i P^i = \beta^1 + (\beta^2 - \beta^1)(P^2 + P^3) + (\beta^3 - \beta^2)P^3$, and hence,

$$D = (\beta^2 - \beta^1)\left(\Psi(P^2+P^3) - (P^2+P^3)\right) + (\beta^3 - \beta^2)\left(\Psi(P^3) - P^3\right). \tag{11}$$

Let $\alpha := P^3$ and $\alpha' := P^2 + P^3$. Since β^3 is a fixed positive number, its probability, P^3, converges to 0 as $\beta^2 \to 0$. Given the assumption $P^2 < \bar{\alpha}$, it follows that for all sufficiently small values of β^2, both α and α' belong to $(0, \bar{\alpha})$, and hence, both $\Psi(\alpha) - \alpha$ and $\Psi(\alpha') - \alpha'$ are positive. By equation (11), this completes the proof of statement (a).

To prove statement (b), suppose now $\beta^2 \geq 1/2$. If α is greater than $\bar{\alpha}$, then so is α'. In this case, $\Psi(\alpha) - \alpha$ and $\Psi(\alpha') - \alpha'$ are both negative, and (11) immediately implies $D < 0$. Thus, without loss of generality let $\alpha \leq \bar{\alpha}$.

Since $\beta^2 \geq 1/2$, we have $\beta^3 - \beta^2 = 1 - \beta^2 \leq \beta^2 = \beta^2 - \beta^1$, and hence, $P^3 \geq P^1$. This implies $P^3 > 1/4$, for we have $P^3 + P^1 = 1 - P^2 > 1 - \bar{\alpha} > 1/2$ by assumptions. As $1/4 > \bar{\alpha}/2$, we then see that $P^3 > \bar{\alpha}/2$. Moreover, $P^3 \geq P^1$ also implies $P^1 \leq 1/2$, and hence, $P^2 + P^3 = 1 - P^1 \geq 1/2 > \bar{\alpha}$. To summarize, α belongs to $(\bar{\alpha}/2, \bar{\alpha}]$ while α' belongs to $(\bar{\alpha}, 1)$.

The next step is to show that

$$\alpha - \bar{\alpha}/2 > |\alpha' - (1+\bar{\alpha})/2|\,. \tag{12}$$

As noted earlier, $\alpha = P^3 \geq P^1 = 1-\alpha'$. Thus, $\alpha - \bar{\alpha}/2 \geq 1-\alpha'-\bar{\alpha}/2 > (1+\bar{\alpha})/2-\alpha'$, where the last inequality follows from the assumption $1/2 > \bar{\alpha}$. Moreover, $P^2 < 1/2$ implies $\alpha' - (1+\bar{\alpha})/2 < \alpha' - (P^2 + \bar{\alpha}/2) = \alpha - \bar{\alpha}/2$.

Inequality (12) and assumption (iv) imply $(\Psi(\alpha') - \alpha') + (\Psi(\alpha) - \alpha) < 0$. Since $\beta^2 - \beta^1 \geq \beta^3 - \beta^2$ and $\Psi(\alpha') - \alpha' < 0 \leq \Psi(\alpha) - \alpha$, from equation (11) it follows that $D < 0$, as we seek. □

Proof of Proposition 4. Since $\alpha > 0$, the conditions in statement (b) ensure that $P^*(\Pi)$ and $\hat{P}^*(\Pi)$ are positive numbers. Set $\lambda := P^*(\mu)/P^*(\Pi)$, $\hat{\lambda} := \hat{P}^*(\mu)/\hat{P}^*(\Pi)$, and $\Pi' := \Pi \backslash \{\mu\}$, where μ stands for $\mu(P^*)$.

By equation (4), we have $\hat{P}^*(\mu) = \alpha + (1-\alpha)P^*(\mu)$. Similarly, $\mu \in \Pi$ implies $\hat{P}^*(\Pi) = \alpha + (1-\alpha)P^*(\Pi)$. Since α and $P^*(\Pi')$ are both positive, it easily follows that $\hat{\lambda} > \lambda$.

Let $\beta := (\hat{\lambda} - \lambda)/(1-\lambda)$. We claim that

$$\hat{P}^*|\Pi = \beta\delta_\mu + (1-\beta)P^*|\Pi. \tag{13}$$

If $\alpha = 1$, then $\beta = 1$ and $\hat{P}^* = \hat{P}^*|\Pi = \delta_\mu$. Suppose now $\alpha < 1$, so that $\hat{P}^*(\Pi') = (1-\alpha)P^*(\Pi') > 0$. In this case, we have $\hat{P}^*|\Pi = \hat{\lambda}\delta_\mu + (1-\hat{\lambda})\hat{P}^*|\Pi'$ and $P^*|\Pi = \lambda\delta_\mu + (1-\lambda)P^*|\Pi'$, while $P^*|\Pi' = \hat{P}^*|\Pi'$ by equation (4). It is a straightforward exercise to derive (13) from these equalities.

Since $\mu_h \succ \mu(P^*|\Pi)_h$, equation (13) and STD(iv) jointly imply $(\hat{P}^*|\Pi)_h = \beta\delta_{\mu_h} + (1-\beta)(P^*|\Pi)_h \succ \beta\delta_{\mu(P^*|\Pi)_h} + (1-\beta)(P^*|\Pi)_h$. If $\succsim$ is strongly compound-risk averse, from (3) we get $(\hat{P}^*|\Pi)_h \succ \beta(P^*|\Pi)_h + (1-\beta)(P^*|\Pi)_h = (P^*|\Pi)_h$. Alternatively, if $\succsim$ is convex on $\Delta^2(X)$ and $\delta_{\mu(P^*|\Pi)_h} \succsim (P^*|\Pi)_h$, we also get $(\hat{P}^*|\Pi)_h \succ \beta\delta_{\mu(P^*|\Pi)_h} + (1-\beta)(P^*|\Pi)_h \succsim (P^*|\Pi)_h$. Indeed, the condition $\mu(P^*|\Pi)_h \succsim_\Pi h$ in statement (c) means $\delta_{\mu(P^*|\Pi)_h} \succsim (P^*|\Pi)_h$. We conclude that $(\hat{P}^*|\Pi)_h \succ (P^*|\Pi)_h$ if (c) or (d) holds.□

References

Abdellaoui, M., Klibanoff, P. and L. Placido (2015): "Experiments on compound risk in relation to simple risk and to ambiguity," *Management Science* **61**, 1306–1322.

Al-Najjar, N.I. and L. De Castro (2014): "Parametric representation of preferences," *Journal of Economic Theory* **150**, 642–667.

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

Arrow, K.J. (1984): Collected Papers of Kenneth J. Arrow, Volume 3: Individual Choice under Certainty and Uncertainty. Cambridge, Massachusetts: Harvard University Press.

Baillon, A., L'Haridon, O. and L. Placido (2011): "Ambiguity models and the Machina paradoxes," *American Economic Review* **101**, 1547–1560.

Briggs, J.S., Cesarini, D., Lindqvist, E. and R. Östling (2021): "Windfall gains and stock market participation," *Journal of Financial Economics* **139**, 57–83.

Cerreia-Vioglio, S., Dillenberger D. and P. Ortoleva (2015): "Cautious expected utility and the certainty effect," *Econometrica* **83**, 693–728.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and L. Montrucchio (2013): "Ambiguity and robust statistics," *Journal of Economic Theory* **148**, 974–1049.

Chateauneuf, A. and M. Cohen (1994): "Risk seeking with diminishing marginal utility in a non-expected utility model," *Journal of Risk and Uncertainty* **9**, 77–91.

Chew, S.H. (1989): "An axiomatic generalization of the quasilinear mean and Gini mean with application to decision theory," Johns Hopkins University, mimeo.

Chew, S.H., Karni, E. and Z. Safra (1987): "Risk aversion in the theory of expected utility with rank dependent probabilities," *Journal of Economic Theory* **42**, 370–381.

Chew, S.H., Miao, B. and S. Zhong (2017): "Partial ambiguity," *Econometrica* **85**, 1239–1260.

Dean, M. and P. Ortoleva (2019): "The empirical relationship between nonstandard economic behaviors," *Proceedings of the National Academy of Sciences* **116**, 16262-16267.

Denti, T. and L. Pomatto (2022): "Model and predictive uncertainty: A foundation for smooth ambiguity preferences," *Econometrica* **90**, 551-584.

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

Dillenberger, D. and U. Segal (2017): "Skewed noise," *Journal of Economic Theory* **169**, 344-364.

Dow, J. and S.R. Werlang (1992): "Uncertainty aversion, risk aversion, and the optimal choice of portfolio," *Econometrica* **60**, 197-204.

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

Ellsberg, D. (2011): "Notes on the origin of the Ellsberg urns (Introduction to the symposium issue)," *Economic Theory* **48**, 221–227.

Epstein, L.G. (2010): "A paradox for the 'smooth ambiguity' model of preference," *Econometrica* **78**, 2085–2099.

Ergin, H. and F. Gul (2009): "A theory of subjective compound lotteries," *Journal of Economic Theory* **144**, 899–929.

Evren, Ö. (2017): "Second-order representations: Anything goes," New Economic School, mimeo. (Available at SSRN: https://papers.ssrn.com/abstract=2941740)

Evren, Ö. (2019): "Recursive non-expected utility: Connecting ambiguity attitudes to risk preferences and the level of ambiguity," *Games and Economic Behavior* **114**, 285–307.

Ghirardato, P. (2002): "Revisiting Savage in a conditional world," *Economic Theory* **20**, 83–92.

Halevy, Y. (2007): "Ellsberg revisited: An experimental study," *Econometrica* **75**, 503–536.

Halevy, Y. and E. Ozdenoren (2022): "Uncertainty and compound lotteries: Calibration," *Economic Theory* **74**, 373–395.

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

Klibanoff, P., Marinacci, M. and S. Mukerji (2012): "On the smooth ambiguity model: A reply," *Econometrica* **80**, 1303–1321.

Klibanoff, P., Mukerji, S., Seo, K. and L. Stanca (2022): "Foundations of ambiguity models under symmetry: α-MEU and smooth ambiguity," *Journal of Economic Theory* **199**, 105202.

Kocher, M.G., Lahno, A.M. and S.T. Trautmann (2018): "Ambiguity aversion is not universal," *European Economic Review* **101**, 268–283.

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

Machina, M.J. (2009): "Risk, ambiguity, and the rank-dependence axioms," *American Economic Review* **99**, 385–392.

Machina, M.J. (2014): "Ambiguity aversion with three or more outcomes," *American Economic Review* **104**, 3814–3840.

Machina, M.J. and D. Schmeidler (1992): "A more robust definition of subjective probability," *Econometrica* **60**, 745–780.

Machina, M.J. and D. Schmeidler (1995): "Bayes without Bernoulli: Simple conditions for probabilistically sophisticated choice," *Journal of Economic Theory* **67**, 106–128.

Nau, R.F. (2006), "Uncertainty aversion with second-order utilities and probabilities," *Management Science* **52**, 136–145.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization* **3**, 323–343.

Schmeidler D. (1989): "Subjective probability and expected utility without additivity," *Econometrica* **57**, 571-587.

Schmidt U. (2004): "Alternatives to expected utility: formal theories," in: Barberà S., Hammond P.J., Seidl C. (eds), Handbook of Utility Theory. Springer, Boston.

Segal, U. (1987): "The Ellsberg paradox and risk aversion: An anticipated utility approach," *International Economic Review* **28**, 175–202.

Segal, U. (1990): "Two stage lotteries without the reduction axiom," *Econometrica* **58**, 349–377.

Seo, K. (2009): "Ambiguity and second-order belief," *Econometrica* **77**, 1575–1605.

Vissing-Jorgensen, A. (2003): "Perspectives on behavioral finance: Does 'irrationality' disappear with wealth? Evidence from expectations and actions," *NBER Macroeconomics Annual*, 139–208.

Wakker, P. (1994): "Separating marginal utility and probabilistic risk aversion," *Theory and Decision* **36**, 1–44.