Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 187 (2020) 105001

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

A theoretical foundation of ambiguity measurement ☆

Yehuda Izhakian

Baruch College, The City University of New York, One Bernard Baruch Way, New York, NY 10010, USA

Received 8 August 2017; final version received 20 January 2020; accepted 27 January 2020
Available online 1 February 2020

Abstract

Ordering alternatives by their degree of ambiguity is crucial in economic and financial decision-making processes. To quantify the degree of ambiguity, this paper introduces an empirically-applicable, outcome-independent (up to a state space partition), risk-independent, and attitude-independent *measure of ambiguity*. In the presence of ambiguity, the Bayesian approach can be extended to uncertain probabilities such that aversion to ambiguity is defined as aversion to mean-preserving spreads in these probabilities. Thereby, the degree of ambiguity can be measured by the volatility of probabilities, just as the degree of risk can be measured by the volatility of outcomes. The applicability of this measure is demonstrated by incorporating ambiguity into an asset pricing model.

JEL classification: D81; D83; G11; G12

Keywords: Ambiguity index; Knightian uncertainty; Ambiguity aversion; Uncertain probabilities; Perceived probabilities; Ambiguity premium

☆ I appreciate helpful comments and suggestions from the editor Laura Veldkamp, an anonymous associate editor, two anonymous referees, Patrick Augustin, Erin Berman, Adam Brandenburger, Menachem Brenner, Xavier Gabaix, Itzhak Gilboa, William Greene, Sergiu Hart, Richard Herron, Zur Izhakian, Sharon Kaminer, Edi Karni, Ruth Kaufman, Ilan Kremer, Dan Levin, Mark Machina, Efe Ok, Jacob Sagi, Thomas Sargent, David Schmeidler, Uzi Segal, Erica Soler-Hampejsek, Sharon Teitelbaum, Paul Wachtel, Peter Wakker, Jan Werner, Jaime Zender, as well as seminar and conference participants at Indiana University, Johns Hopkins University, New York University, Ohio State University, The Hebrew University of Jerusalem, University of Colorado, University of Houston, University of Michigan, University of Minnesota, University of Montreal, University of North Carolina, Decision: Theory, Experiments and Applications (D-TEA), Foundations of Utility and Risk (FUR), Risk Uncertainty and Decision (RUD), and University of Chicago Workshop on Ambiguity and Robustness in Macroeconomics and Finance. I acknowledge hosting by the Stern School of Business, New York University.

E-mail address: yehuda.izhakian@baruch.cuny.edu.

https://doi.org/10.1016/j.jet.2020.105001
0022-0531/

> "Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated."
>
> [— Frank Hyneman Knight, 1921]

1. Introduction

Almost every economic and financial decision entails ambiguity (Knightian uncertainty). Ordering alternatives by their degree of ambiguity is crucial for making optimal decisions. Such an ordering requires a well-defined measure of the degree of ambiguity. Knight (1921) recognized that the prerequisite in any theoretical and empirical study of the distinct implications of ambiguity is to measure this dimension of uncertainty, ultimately independent of risk and of attitudes toward ambiguity and risk.[1] A theoretical foundation of a risk- and attitude-independent ambiguity measurement that addresses this need is not available in the literature.

This paper proposes an applicable *ambiguity measure* that is risk independent, simple, and intuitive; it measures the degree of ambiguity independently of individuals' attitudes (tastes) and can be computed from the data. These are key qualities for the introduction of ambiguity into economic and financial models. These qualities are necessary for any study investigating the effect of ambiguity without confounding it with risk or with attitudes toward ambiguity and risk. In theoretical studies, a risk- and attitude-independent ambiguity measurement is necessary for tracking the impact of ambiguity in isolation of risk and individuals' attitudes in comparative statics of models. In empirical and behavioral studies, the aforementioned measurement is necessary for identifying and analyzing the effect of ambiguity on human behavior and decisions, and on economic and financial phenomena and anomalies.[2] The aforementioned measurement is as necessary for the examination of the impact of ambiguity on decision making as is the measurement of risk (independent of attitudes toward risk) for the examination of its impact on decision making. Other factors that have been taken as a proxy for ambiguity are either risk dependent,[3] or cannot be computed from the data,[4] mainly because they are confounded with attitudes toward ambiguity.

To measure the degree of ambiguity independent of risk, the measure must be *outcome independent* up to a state space partition. That is, if the outcomes associated with events change, while the induced partition of the state space into events remains unchanged, then the degree of ambiguity remains unchanged, since all probabilities remain unchanged. In other words, the measurement applies exclusively to the probabilities of events, independent of their outcomes. To illustrate, consider an asset whose payoff is determined by a flip of an unbalanced coin, for which the decision maker (DM) does not know the odds of heads or tails. The payoff of the asset is \$100 in the case of heads and \$0 in the case of tails. Suppose that prior to the coin being

---

[1] Risk is defined as a condition in which the event to be realized is a priori unknown, but the odds of all possible events are perfectly known. Ambiguity refers to conditions in which not only is the event to be realized a priori unknown, but the odds of events are also either not uniquely assigned or are unknown.

[2] In behavioral studies, for example, a risk-independent ambiguity measurement allows for answering questions such as whether attitudes toward ambiguity and risk change with the extent of ambiguity. In empirical studies, for example, such a measurement allows for the analysis of the distinct effect of ambiguity on pricing assets and optimal portfolio selection, and the assessment of the extent of ambiguity associated with a particular stock or stock market, independent of the associated risk, as well as the investors' aversion to ambiguity and risk.

[3] For example, the variance of the mean, the variance of the variance, and the disagreement among analysts' forecasts.

[4] For example, the sum of nonadditive probabilities, the discrimination among different levels of likelihood, matching probabilities, and the difference between the minimal possible mean and the true mean.

flipped, the payoff in the case of heads is suddenly changed to \$1,000; since no new information about probabilities has been obtained, the DM has no reason to change the assessed probabilities or the assessed degree of ambiguity.[5] Therefore, ambiguity is *outcome independent*. Clearly, in this case, risk increases, since risk is *outcome dependent*.

To derive such a theoretical outcome-independent measure of ambiguity, in the underlying decision-making model, the preferences for ambiguity must be outcome independent. Namely, the preferences for ambiguity must apply exclusively to the probabilities of events, independently of the outcomes associated with these events. In addition, the decision-making model must separate attitudes (tastes) from beliefs with respect to ambiguity. Without these two properties, ambiguity cannot be measured independently of risk and of attitudes toward ambiguity and risk. This paper employs the expected utility with uncertain probabilities (EUUP) model, proposed in Izhakian (2017), since it satisfies these two prerequisites. While making a significant contribution to the literature, in earlier models of decision making under ambiguity, either beliefs are not separated from attitudes with respect to ambiguity (e.g., Gilboa and Schmeidler, 1989; Schmeidler, 1989; Tversky and Kahneman, 1992), or the preferences for ambiguity are outcome dependent (e.g., Klibanoff et al., 2005; Nau, 2006; Chew and Sagi, 2008). Moreover, different classes of preferences do not always arrive at the same decision rule and might provide different orderings by ambiguity. Therefore, a specific class of preferences must be selected.[6]

The main idea of the EUUP model is that perceived probabilities, used to assess expected utility, are formed by the "certainty equivalent probabilities" of uncertain probabilities. An uncertain probability is modeled explicitly in a state space that is subject to a prior probability. A perceived probability is the unique certain probability value that the DM is willing to accept in exchange for the uncertain probability of a given event, based on her preference for ambiguity. This preference applies exclusively to probabilities, such that the attitude toward ambiguity is outcome independent and defined as an attitude toward mean-preserving spreads in *probabilities*, analogous to the Rothschild and Stiglitz (1970) attitude toward mean-preserving spreads in *outcomes*.[7,8] Thereby, their approach can be used for uncertain probabilities to define an ordering by ambiguity.

The current paper shows that, by ordering uncertain probabilities, the degree of ambiguity can be measured by the expected *volatility of probabilities*, across the relevant events. Formally, the measure of ambiguity is given by

$$\mho^2[f] \equiv \int_X \mathrm{E}\left[\varphi_f(x)\right] \mathrm{Var}\left[\varphi_f(x)\right] dx, \tag{1}$$

where f is an act (a mapping from states into consequences); X is a convex subset (an interval) of the real line (consequences); $\varphi_f(\cdot)$ is an uncertain probability density function; and the ex-

[5] If, however, in the case of heads the payoff decreases to \$0, then the induced partition of the state space changes and, therefore, the degree of ambiguity changes.

[6] Models of decision making under ambiguity are "seemingly different (...) rarely related to one another, and often expressed in drastically different formal languages" (Epstein and Schneider, 2010, p. 316).

[7] Unlike in the EUUP model, Klibanoff et al. (2005), Nau (2006), and Chew and Sagi (2008) define attitude toward ambiguity as attitude toward mean-preserving spreads in expected (or certainty equivalent) utilities, which are outcome dependent and, therefore, subject to risk and attitude toward risk.

[8] Rothschild and Stiglitz (1972) point out that various components of the key theorem in their 1970 paper could have been drawn from earlier mathematics of Hardy et al. (1953), and Blackwell and Girshick (1954).

pectation $\mathrm{E}[\cdot]$ and the variance $\mathrm{Var}[\cdot]$ are taken using a second-order probability measure ξ on a set $\mathcal{P}$ of probability measures.[9]

The measure $\mho^2$ (mho^2) quantifies the degree of ambiguity associated with beliefs in isolation of attitudes toward ambiguity and risk. The main advantage of $\mho^2$ is that it can be computed from the data and employed in empirical studies (e.g., Izhakian and Yermack, 2017; Brenner and Izhakian, 2018). *Risk independence* is another major advantage of $\mho^2$. Given a fixed set of probability measures on the state space, $\mho^2$ does not depend on the magnitude of the outcomes associated with events, but only on the probabilities of these events, which are defined by the partition that the outcomes induce (through measurable functions—acts) over the state space (i.e., outcome independent up to a state space partition).[10] Thereby, the degree of ambiguity is measured independently of the degree of risk.

Prior studies mainly focus on aversion to ambiguity rather than on quantifying ambiguity (e.g., Bossaerts et al., 2010). Yet, several approaches to estimating ambiguity have been proposed. Dow and Werlang (1992) measure ambiguity by the sum of the nonadditive probabilities. Baillon et al. (2018) extend this approach by measuring discrimination among different levels of likelihood. Baillon and Bleichrodt (2015) propose five ambiguity indices using matching non-additive probabilities. These measures, however, confound ambiguity attitudes and cannot be computed from the data. Ui (2011) measures ambiguity by the difference between the minimal possible mean and the true mean. Bewley (2011) and Boyle et al. (2011) measure ambiguity as a critical confidence interval. Maccheroni et al. (2013) measure ambiguity by the variance of an unknown mean. These studies assume that the variance of outcomes is known and suggest risk-dependent measures, based only on the variation of the mean.[11] However, ambiguous variance has been shown to be an important element in decision making processes (e.g., Epstein and Ji, 2013, 2014).[12]

Unlike in prior studies, the ambiguity measure $\mho^2$ is outcome independent (up to a state space partition), risk independent, and conceptually encompasses both ambiguous mean and ambiguous variance, as well as the ambiguity of all higher moments of the probability distribution (i.e., skewness, kurtosis, etc.), through the uncertainty of probabilities. Relative entropy (Kullback-Leibler divergence), measured by the deviation of a single prior from a reference probability distribution (reference model), may also be interpreted as a risk-independent measure of ambiguity (e.g., Hansen et al., 1999; Hansen and Sargent, 2001). However, while relative entropy is an asymmetric measure of the distance of a single prior relative to a reference probability dis-

[9] In a finite state space, $\mho^2[f] \equiv \sum_j \mathrm{E}\left[\varphi_f\left(x_j\right)\right] \mathrm{Var}\left[\varphi_f\left(x_j\right)\right]$, where $\varphi_f(\cdot)$ is an uncertain probability mass function.

[10] This implies that acts that induce the same partition of the state space into events have the same degree of ambiguity, even though they may associate different outcomes to events and, therefore, may have different degrees of risk. In other words, given a set of events, increasing or decreasing the outcomes associated with each event does not change the degree of ambiguity (as long as both the partition of the state space into events induced by the outcomes and the set of probability measures on the state space remain unchanged), but it might change the degree of risk.

[11] In particular, these measures are sensitive to outcome multiplication by a non-zero constant and are, therefore, outcome- and risk dependent. The same holds true for the variance of the variance (e.g., Faria and Correia-da Silva, 2014) and for the disagreement among analysts' forecasts (e.g., Anderson et al., 2009), which are occasionally interpreted as measures of ambiguity.

[12] In empirical asset pricing and macroeconomic contexts, stochastic volatility also plays an important role (e.g., Bollerslev et al., 1988; Fernández-Villaverde et al., 2010; Bollerslev et al., 2011).

tribution, $\mho^2$ is a symmetric measure that can be applied to multiple priors when a reference distribution cannot be identified.

A practical, risk- and attitude-independent ambiguity measure is an important instrument for studying (both theoretically and empirically) the distinct role that the extent of ambiguity plays in decision-making processes. There are many applications where such a measure can be utilized. The practicality of $\mho^2$ in theoretical studies is demonstrated, for example, in modeling the tradeoff theory of capital structure (Modigliani and Miller, 1958) under ambiguity (Izhakian et al., 2017); pricing of credit default swaps (CDSs) in general equilibrium (Augustin and Izhakian, 2019); and contract efficiency subject to the principal (or the agent) inducing a specific partition of the state space into events (Izhakian and Zender, 2018). The practicality of $\mho^2$ in empirical studies is demonstrated, for example, in analyzing the role ambiguity plays in employees' decisions to exercise their options (Izhakian and Yermack, 2017); the effect of ambiguity on the market equity premium (Brenner and Izhakian, 2018); the effect of ambiguity on firms' payout policy (Herron and Izhakian, 2018); and the effect of ambiguity on trading activities around macroeconomic news (Doan et al., 2018).[13]

To further demonstrate the applicability of $\mho^2$, this paper generalizes the Arrow-Pratt asset pricing theory to incorporate ambiguity, and shows that the price of an asset is determined not only by its degree of risk and the attitude toward risk, but also by its degree of ambiguity and the attitude toward ambiguity. The paper constructs an ambiguity premium, attributed to ambiguity and attitude toward ambiguity, completely distinguished from the risk premium. Empirically, this premium has been shown to significantly affect asset prices (Brenner and Izhakian, 2018).

The paper proceeds as follows. Section 2 presents the decision-making framework, and Section 3 simplifies it. Section 4 develops ordering of events by the degree of ambiguity. Section 5 furthers this ordering to propose a measure of ambiguity. Section 6 analyzes the properties of the proposed measure and compares it to alternative measures. To demonstrate usability, Section 7 models the ambiguity premium and advises how to compute the ambiguity measure from the data. Section 8 concludes.

2. The decision-making framework

In the EUUP decision-making framework, there are two tiers of uncertainty: one with respect to consequences (outcomes), and the other with respect to the probabilities of these consequences. A DM in this framework applies two differentiated phases of the decision process, each reflecting one of these tiers. In the first phase, she forms a representation of her perceived probabilities for all the events relevant to her decision, as the certainty equivalent probabilities of the uncertain probabilities. In the second phase, she assesses the value of each alternative using her perceived probabilities and chooses accordingly. Ambiguity—the uncertainty about probabilities—plays a key role in the first phase, while risk—the uncertainty about consequences—plays a key role in the second phase.

[13] Accounting for ambiguity might provide insight on some other phenomena that previously could not have been fully explained. Notable examples include the tendency to hold very small portfolios (Goetzmann and Kumar, 2008); the risk-free rate puzzle (Weil, 1989); the limited ability of changes in the fundamental to justify the observed equity volatility (Shiller, 1981); and the home bias puzzle (Coval and Moskowitz, 1999).

Let $\mathcal{S}$ be an infinite *primary* state space, endowed with a σ-algebra, $\mathcal{E}$, of subsets of $\mathcal{S}$.[14] Generic elements of this σ-algebra are called *events* and are denoted by E. A λ-system $\mathcal{H} \subset \mathcal{E}$ should be thought of as containing events with an unambiguous probability (i.e., events with a known, objective probability).[15] Define $X \subseteq \mathbb{R}$ to be an interval (a convex set) of *consequences* in $\mathbb{R}$ that contains the interval $[0, 1]$. A consequence $x \in X$ is considered to be unfavorable if $x \le k$ and favorable if $k < x$, where k is a *reference point*.[16] Let a *primary act* $f : \mathcal{S} \to X$ be a bounded $\mathcal{E}$-measurable function from states into consequences, and denote the set of all these simple measurable (Savage) acts by $\mathcal{F}_0$.[17] A *primary indicator act*, $\delta_E = \left(E^C : 0, E : 1\right)$, assigns the outcome 1 to event $E \in \mathcal{E}$ and the outcome 0 to its complementary event $E^C \in \mathcal{E}$. Denote by $\succsim^1$ the DM's preference relation over the set of simple primary acts $\mathcal{F}_0$, and let the relations $\precsim^1$, $\prec^1$, $\succ^1$, and $\sim^1$ be defined as usual.

The probabilities of events $E \in \mathcal{E}$ occurring in the primary space are determined in a nonempty *secondary* space, defined by a set $\mathcal{P}$ of all additive probability measures on the primary space $\mathcal{S}$ that agree on $\mathcal{H}$. A *first-order probability measure* $\mathrm{P} \in \mathcal{P}$ is then viewed as a state of nature in this secondary space, and the state space $\mathcal{P}$ is assumed to be endowed with an algebra, $\Pi \subset 2^{\mathcal{P}}$, of subsets of $\mathcal{P}$ that satisfy the structure required by Kopylov (2010). A *secondary act*, $\widehat{f} : \mathcal{P} \to X$, is a bounded Π-measurable function from the secondary space $\mathcal{P}$ into the set of consequences X. The set of all secondary acts is denoted $\widehat{\mathcal{F}}$. A secondary act $\hat{\delta}_E : \mathcal{P} \to [0, 1]$ that describes the resulting expected outcome of a (primary) indicator act $\delta_E \in \mathcal{F}_0$ is given by

$$\hat{\delta}_E\left(\mathrm{P}\right) = \mathrm{P}\left(E\right),$$

for every $\mathrm{P} \in \mathcal{P}$, and the subset of all such $\hat{\delta}_E \in \widehat{\mathcal{F}}$ is denoted $\hat{\Delta}$. A secondary act $\hat{\delta}_E$ can, therefore, be described as a function that assigns each event $E \in \mathcal{E}$ with its possible probabilities. In this view, $\hat{\delta}_E$ can be interpreted as an uncertain variable describing the probability $\{\mathrm{P}\left(E\right)\}_{\mathrm{P}\in\mathcal{P}}$ of event E.

In view of $\hat{\delta}_E \in \hat{\Delta}$ as describing the (uncertain) probability of event E, the preference relation $\succsim^2$ over $\hat{\Delta}$ may well be referred to as a preference over probabilities.[18] To illustrate, consider the two secondary acts $\hat{\delta}_E, \hat{\delta}_F \in \hat{\Delta}$, associated with the indicator acts $\delta_E, \delta_F \in \mathcal{F}_0$, and assume a DM who prefers $\hat{\delta}_E$ to $\hat{\delta}_F$. This means that she prefers the good outcome with the (uncertain) probability $\{\mathrm{P}\left(E\right)\}_{\mathrm{P}\in\mathcal{P}}$ than with the (uncertain) probability $\{\mathrm{P}\left(F\right)\}_{\mathrm{P}\in\mathcal{P}}$. Therefore, secondary acts in $\hat{\Delta}$ can be considered as bets on the probabilities of events. In the EUUP model, the preference relation $\succsim^2$ is utilized solely to identify the DM's certainty equivalent probability (perceived probability) of each event $E \in \mathcal{E}$ separately. Therefore, $\succsim^2$ over only secondary acts that are associated with primary indicator acts, $\hat{\Delta}$, play a role in the decision making. In the EUUP model,

[14] Several notational conventions are used: constants and variables are italicized, operators in regular font (followed by square parentheses), and sets in capital calligraphic font.

[15] The set $\mathcal{H}$ may not necessarily be an algebra, as it may not be closed with respect to intersections. Yet, a probability measure can be defined on $\mathcal{H}$ (Epstein and Zhang, 2001). These settings apply to any case where there is a sub σ-algebra isomorphic to the Borel sets on $[0, 1]$ endowed with Lebesgue measure. In this case, the extended state space can be defined by $\mathcal{S} \times [0, 1]$ (e.g., Sarin and Wakker, 1992).

[16] The reference point, k, also referred to as the status quo, is subjectively defined by each DM and may be affected by the formulation of the offered prospects (e.g., Kahneman and Tversky, 1979).

[17] A simple primary act can be represented as a sequence of pairs, $f = (E_1 : x_1, \ldots, E_n : x_n)$, where $(E_1, \ldots, E_n)$ is a generic partition of the state space $\mathcal{S}$; x_j is the consequence if event E_j occurs; and the consequences $x_1, \ldots, x_n$ are listed in a non-decreasing order. To support primary acts taking infinitely many values, $\mathcal{F}_0$ may be replaced by a set of all bounded measurable primary acts that includes all simple acts and satisfies the structure of Kothiyal et al. (2011).

[18] Section 4 elaborates on this interpretation.

$\succsim^2$ implies a unique second-order countably-additive probability measure ξ on Π that assigns each subset $A \in \Pi$ with a probability $\xi(A)$.

Under the EUUP assumptions (detailed in Appendix A.1), there exists a function $\mathrm{V} : \mathcal{F}_0 \to \mathbb{R}$ such that

$$f \succsim^1 g \quad \Longleftrightarrow \quad \mathrm{V}(f) \geq \mathrm{V}(g),$$

for every $f, g \in \mathcal{F}_0$, where

$$\mathrm{V}(f) = \int_{z \leq 0} \left[\Upsilon^{-1} \left(\int_{\mathcal{P}} \Upsilon\big(\mathrm{P}(\{s \in \mathcal{S} \,|\, \mathrm{U}(f(s)) \geq z\}) \big) d\xi \right) - 1 \right] dz + \tag{2}$$
$$\int_{z \geq 0} \Upsilon^{-1} \left(\int_{\mathcal{P}} \Upsilon\big(\mathrm{P}(\{s \in \mathcal{S} \,|\, \mathrm{U}(f(s)) \geq z\}) \big) d\xi \right) dz;$$

$\mathrm{U} : X \to \mathbb{R}$ is a continuous, strictly increasing bounded function, normalized such that $\mathrm{U}(k = 0) = 0$[19]; $\Upsilon : [0, 1] \to \mathbb{R}$ is a continuous, non-constant bounded function; ξ is a non-atomic, countably-additive probability measure; and every $\mathrm{P} \in \mathcal{P}$ is a finitely-additive probability measure. Furthermore, ξ is unique, Υ is unique up to a positive linear transformation, and U is unique up to a positive scaling.[20]

The function V, representing the DM's preferences, takes the form of a reference-dependent Choquet integration over unfavorable and favorable outcomes (relative to the reference point k). It adds to the Choquet expected utility theory (CEU, Schmeidler, 1989) and cumulative prospect theory (CPT, Tversky and Kahneman, 1992) a construction of capacities from beliefs and attitudes toward ambiguity. V distinguishes between beliefs and attitudes (tastes), and between ambiguity and risk. First-order beliefs, which capture risk, are formed by the uncertain probability measure P. Second-order beliefs, which capture ambiguity, are formed by the probability measure ξ on a set of priors $\mathcal{P}$ on $\mathcal{S}$. Attitudes toward risk are formed by the utility function U. Attitudes toward ambiguity are formed by the function Υ, referred to as an *outlook function*. This function can be thought of as if the DM uses it to form her certainty equivalent probabilities of each event separately. As with risk attitudes, there are three types of attitudes toward ambiguity: ambiguity aversion (formed by a concave Υ), ambiguity seeking (formed by a convex Υ), and ambiguity indifference (formed by a linear Υ). The EUUP framework allows different attitudes toward ambiguity and risk concerning unfavorable and favorable outcomes. However, for simplicity, it is assumed that the DM has symmetric preferences for ambiguity and risk.

Technically, in most cases, (secondary) acts are not observable and beliefs are not verifiable, even in the frameworks of classical decision-making models that do not consider ambiguity. For example, to identify beliefs in the Savage framework, the DM has to be presented with infinitely many choices. In the max-min expected utility with multiple priors framework (Gilboa and Schmeidler, 1989), the DM acts as if she chooses from sets of priors, while the sets of

[19] The reference point (the "status quo") k is assumed to be 0.

[20] Superficially, the model proposed in Equation (2) resembles the smooth model (Klibanoff et al., 2005). However, although both models rely on the idea of second-order preferences (preferences for ambiguity) over secondary acts, they are conceptually different. Preferences for ambiguity in the smooth model are outcome dependent, whereas in the EUUP model they are outcome independent. The axiomatic differences between the smooth model and the EUUP model are discussed in Appendix A.1.

(first-order) priors are not always identifiable and the realized prior is not verifiable. In model uncertainty (e.g., Hansen et al., 1999; Hansen and Sargent, 2001), the reference (true) distribution is unidentifiable. Similarly, the use of secondary acts might be challenged by the claim that these acts are imaginary objects and a DM cannot bet on a set of probability measures, as the realized probabilities are not verifiable. The use of secondary acts hinges on Klibanoff et al. (2005), who introduce the concept of secondary acts.[21] Implicitly or explicitly, many other studies assume second-order acts or second-order beliefs (e.g., Nau, 2006; Chew and Sagi, 2008; Seo, 2009; Hayashi and Miao, 2011). Some earlier formal decision theory studies also rely on similarly unobservable acts (e.g., Karni and Schmeidler, 1981; Karni et al., 1983; Grant and Karni, 2004). Previous behavioral studies indicate that DMs act as if they have second-order preferences (e.g., Ellsberg, 1961; Halevy, 2007; Hao and Houser, 2012). Furthermore, neural responses to second-order uncertainty (ambiguity) are associated with different brain areas than those supporting first-order uncertainty (e.g., Huettel et al., 2006; Bach et al., 2011).

3. Preliminaries

To elicit a measure of the extent of ambiguity, the functional representation V has to be simplified. The key for simplification is the DM's perceived probabilities (capacities), which underpin the DM's preferences representation in Equation (2). Given an event, the perceived probability, derived from the uncertainty about probabilities (ambiguity) and from the DM's attitude toward this uncertainty, is formed as the certainty equivalent probability of the event's uncertain probability. Namely, the perceived probability is the unique certain probability value that the DM is willing to accept in exchange for the uncertain probability of a given event. Formally, the *perceived probability*

$$\mathrm{Q}(E) = \Upsilon^{-1}\left(\int_{\mathcal{P}} \Upsilon\left(\mathrm{P}\left(E\right)\right) d\xi\right) \tag{3}$$

of event $E \in \mathcal{E}$ is obtained by applying subjective expected utility axioms (e.g., Kopylov, 2010) to the probabilities of E.[22] Notice that perceived probabilities are independent of the magnitude of outcomes assigned by a primary act f to states of nature in the state space $\mathcal{S}$, but are dependent on the partition that f induces on $\mathcal{S}$ into events. This representation is possible since, in the EUUP model, the preferences for ambiguity, $\succsim^2$, are outcome independent (i.e., they apply to the probabilities of events independently of their associated outcomes).

Equation (3) proposes that the DM, who views uncertain probabilities as a set of priors, aggregates these probabilities through the outlook function Υ in a nonlinear way to form her perceived probabilities. As a consequence of probabilistic sensitivity (i.e., the nonlinear ways in which individuals may interpret probabilities), the perceived probabilities are nonadditive. Ambiguity aversion (a concave outlook function) results in a subadditive probability measure, while ambiguity seeking (a convex outlook function) results in a superadditive measure. As in

[21] In particular, Klibanoff et al. (2005) stress that "second order acts are not as strange or unfamiliar as they might first appear. (...) In a parametric portfolio investment example, these could be bets about the parameter values that characterize the asset returns, e.g., means, variances, and covariances. Similarly, in model uncertainty applications, second order acts are bets about the values of the relevant parameters in the underlying model. Closer to decision theory, for an Ellsberg urn, second order acts may be viewed as bets on the composition of the urn."

[22] See Proposition 2 in Izhakian (2017).

CEU (Schmeidler, 1989) and CPT (Tversky and Kahneman, 1992), the perceived probabilities (capacities) in the EUUP are endogenously determined.

To write the dual representation of the value function V, the following notation is used: $\mathrm{P}_f(x)$ denotes the cumulative probability $\mathrm{P}\big(\{s \in \mathcal{S} \mid f(s) \leq x\}\big)$, and $\varphi_f(x)$ denotes the marginal (mass) probability $\mathrm{P}\big(\{s \in \mathcal{S} \mid f(s) = x\}\big)$.[23] Suppose that $\varphi_f(x)$ exists, bounded and well defined for every $x \in X$. By Wakker and Tversky (Equation 6.1, 1993), the dual representation $\mathrm{W}: \mathcal{F}_0 \to \mathbb{R}$ of $\mathrm{V}: \mathcal{F}_0 \to \mathbb{R}$ can then be written as

$$\mathrm{W}(f) = -\int\limits_{x \leq k} \mathrm{U}(x)\, d\left[\Upsilon^{-1}\left(\int\limits_{\mathcal{P}} \Upsilon\big(1 - \mathrm{P}_f(x)\big)\, d\xi\right) - 1\right] + \int\limits_{x \geq k} \mathrm{U}(x)\, d\left[\Upsilon^{-1}\left(\int\limits_{\mathcal{P}} \Upsilon\big(1 - \mathrm{P}_f(x)\big)\, d\xi\right)\right], \tag{4}$$

where $d\left[\Upsilon^{-1}\left(\int\limits_{\mathcal{P}} \Upsilon\big(1 - \mathrm{P}_f(x)\big)\, d\xi\right)\right]$ is the marginal perceived probability derived from Equation (3).

The marginal perceived probabilities can be simplified by utilizing a Taylor expansion. To this end, the *expected* cumulative and marginal probability of $x \in X$ are, respectively, defined by[24]

$$\mathrm{E}\left[\mathrm{P}_f(x)\right] \equiv \int\limits_{\mathcal{P}} \mathrm{P}_f(x)\, d\xi \quad \text{and} \quad \mathrm{E}\left[\varphi_f(x)\right] \equiv \int\limits_{\mathcal{P}} \varphi_f(x)\, d\xi,$$

and the *variance* of the marginal probability by

$$\mathrm{Var}\left[\varphi_f(x)\right] \equiv \int\limits_{\mathcal{P}} \big(\varphi_f(x) - \mathrm{E}\left[\varphi_f(x)\right]\big)^2 d\xi.$$

The expected outcome and the variance of outcomes are defined, respectively, by the double expectation (with respect to probabilities and to outcomes)[25]:

$$\mathbb{E}[f] \equiv \int\limits_{X} \mathrm{E}\left[\varphi_f(x)\right] x\, dx \quad \text{and} \quad \mathbb{V}\mathrm{ar}[f] \equiv \int\limits_{X} \mathrm{E}\left[\varphi_f(x)\right] \big(x - \mathbb{E}[f]\big)^2 dx.$$

When clear from the context, the subscript f, denoting an act, is omitted. A double-struck capital font designates expectation or variance of outcomes with respect to expected probabilities, while regular font designates expectation or variance of probabilities with respect to second-order probabilities.

[23] In the case of acts taking infinitely many values, the marginal (density) probability can be defined by $\mathrm{P}\big(\{s \in \mathcal{S} \mid x - \epsilon < f(s) < x + \epsilon\}\big)$, where ϵ is small enough (e.g., Papoulis and Pillai, 1965, p. 32).

[24] Notice that the operator $\mathrm{E}[\cdot]$ (in straight font followed by square parentheses) denotes expectation, while the variable E (in slanted italic font) denotes an event in $\mathcal{E}$.

[25] Specifically, $\mathbb{E}[f] \equiv \int\limits_{\mathcal{P}} \int\limits_{\mathcal{S}} f\, dP\, d\xi$ and $\mathbb{V}\mathrm{ar}[f] \equiv \int\limits_{\mathcal{P}} \int\limits_{\mathcal{S}} f^2 dP\, d\xi - \left(\int\limits_{\mathcal{P}} \int\limits_{\mathcal{S}} f\, dP\, d\xi\right)^2$.

Theorem 1. *Suppose that Υ is strictly-increasing, twice-differentiable, and satisfies*

$$\left| \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} \right| \leq \frac{1}{\mathrm{Var}\left[\varphi_f(x)\right]}.$$

The marginal perceived probability of x can then be written as

$$\frac{d}{dx}\left[\Upsilon^{-1}\left(\int_{\mathcal{P}} \Upsilon\left(1-\mathrm{P}_f(x)\right) d\xi\right)\right] \tag{5}$$
$$= \begin{cases} \mathrm{E}\left[\varphi_f(x)\right]\left(1-\dfrac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}\mathrm{Var}\left[\varphi_f(x)\right]\right)+R_2\left(\varphi_f(x)\right), & \text{if } x \leq k, \\ \mathrm{E}\left[\varphi_f(x)\right]\left(1+\dfrac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}\mathrm{Var}\left[\varphi_f(x)\right]\right)+R_2\left(\varphi_f(x)\right), & \text{if } x > k. \end{cases}$$

Moreover, the remainder $R_2\left(\varphi_f(x)\right) = o\left(\mathrm{E}\left[\left|\varphi_f(x)-\mathrm{E}\left[\varphi_f(x)\right]\right|^3\right]\right)$ as $\left|\varphi_f(x)-\mathrm{E}\left[\varphi_f(x)\right]\right| \to 0$.[26]

Theorem 1 establishes a local expansion of the marginal perceived probabilities.[27] The remainder of order $o\left(\mathrm{E}\left[\left|\varphi_f(x)-\mathrm{E}\left[\varphi_f(x)\right]\right|^3\right]\right)$ means that the fourth and higher absolute central moments of the uncertain marginal probability $\varphi_f(x)$ are of a strictly smaller order than the third absolute central moment of $\varphi_f(x)$, and are therefore negligible. The remainder is identically zero—the expansion in Equation (5) is exact—when Υ is quadratic or linear, or when the fourth and higher absolute central moments of $\varphi_f(x)$ are equal to zero. The condition on Υ bounds the level of ambiguity aversion (the concavity of Υ) and the level of ambiguity seeking (the convexity of Υ) to ensure that the approximated marginal perceived probabilities are nonnegative.[28] Note that Theorem 1 assumes no uniform conditions over primary or secondary acts. It applies to the perceived probabilities of every $x \in X$ under any $f \in \mathcal{F}_0$ with any uncertain probability $\hat{\delta}_{\{s\in\mathcal{S} \mid f(s)=x\}} \in \hat{\Delta}$.

Theorem 2. *Suppose that the conditions of Theorem 1 are satisfied. The dual representation $\mathrm{W} : \mathcal{F}_0 \to \mathbb{R}$ of the DM's preferences can then be written as*[29]

$$\mathrm{W}(f) = \int_{x\leq k} \mathrm{U}(x)\,\mathrm{E}\left[\varphi_f(x)\right]\left(1-\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}\mathrm{Var}\left[\varphi_f(x)\right]\right) dx + \tag{6}$$

26 The notation of the remainder, R_2, is relative to the nth derivative of Υ.

27 A similar approach is taken by Pratt (1964) and Arrow (1965), who apply a quadratic expansion to expected utility with respect to uncertain outcomes.

28 This implies that the approximated perceived probabilities satisfy $\mathrm{Q}(\emptyset)=0$, $\mathrm{Q}(\mathcal{S})=1$, and set monotonicity with respect to set-inclusion, i.e., $\mathrm{Q}(E) \leq \mathrm{Q}(F)$ if $E \subset F$.

29 Recall that k is the DM's subjective reference point, relative to which consequences are classified as unfavorable or favorable. Referring to the violation of expected utility, Wakker and Tversky (1993, p. 147) describe the idea of a reference point as follows: "One notion is that the objects of choice are prospects, defined in terms of gains and losses relative to neutral reference point, rather than acts defined in terms of final asset positions. (...) The significance of the reference point stems from the observations that people are generally risk averse for gains, risk seeking for losses, and that losses loom larger than gains."

$$\int_{x\geq k} \mathrm{U}(x)\,\mathrm{E}\left[\varphi_f(x)\right]\left(1+\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}\mathrm{Var}\left[\varphi_f(x)\right]\right)dx+R_2(f).$$

That is, for every $f, g \in \mathcal{F}_0$,

$$f \succsim^1 g \iff \mathrm{W}(f) \geq \mathrm{W}(g)\,.$$

Moreover, the remainder $R_2(f) = o\left(\int \mathrm{E}\left[\left|\varphi_f(x)-\mathrm{E}\left[\varphi_f(x)\right]\right|^3\right] x\,dx\right)$ *as* $\int \left|\varphi_f(x) - \mathrm{E}\left[\varphi_f(x)\right]\right| dx \to 0$.

The representation of $\succsim^1$, established in Theorem 2, is based on the simplified marginal perceived probabilities, developed in Theorem 1. Hence, the order of the remainder (the "error") in Equation (6) is determined by the order of the remainder in Equation (5). Therefore, the order of the error is three times smaller than the order of the outcomes in X and tends to zero as $\int \left|\varphi_f(x) - \mathrm{E}\left[\varphi_f(x)\right]\right| dx \to 0$, i.e., as the deviation of probabilities from their expectations tends to zero. This is equivalent to a cubic expansion, i.e., remainder of order $o\left(\mathbb{E}\left[|x-\mathbb{E}[x]|^3\right]\right)$, in which the fourth and higher absolute central moments of outcomes in X are of a strictly smaller order than the third absolute central moment as $|x-\mathbb{E}[x]| \to 0$, and are therefore negligible.

Since the expansion is applied only to marginal perceived probabilities, which are solely a function of probabilities and attitudes toward their spreads, Theorem 2 makes assumptions neither about primary acts and their outcomes nor about the utility function U. Similarly to Theorem 1, to expand W, no uniform conditions over secondary acts are made. Notably, the bounding condition on Υ, inherited from Theorem 1, is required only for the expansion of W, not for the representation of preferences by W, defined in Equation (4).

4. Ordering ambiguous events

To order primary acts (bets) by the degree of ambiguity, ambiguous events have to first be identified, then an ordering by the degree of ambiguity over these events has to be defined.

Definition 1. An event $E \in \mathcal{E}$ is unambiguous if and only if its probability is agreed upon by all probability measures in $\mathcal{P}$, i.e., $E \in \mathcal{H}$.

Definition 1 implies that an event is unambiguous if and only if its probability is uniquely defined, and is known and objective. It relies only on beliefs, independent of attitudes toward ambiguity. Definition 1 coincides with Chen and Epstein (2002) and Ghirardato and Marinacci (2002), who define E as unambiguous if the probabilities of E and its complementary E^C are additive. Klibanoff et al. (2005) define E as ambiguous if betting on E is less desirable than betting on some event F, and betting on E^C is also less desirable than betting on F^C. Since their definition relies on attitudes toward ambiguity, the same set of unambiguous events as in Definition 1 is obtained only when attitudes toward ambiguity are symmetric. Epstein and Zhang (2001) provide a definition of unambiguous events that applies to a wide range of ambiguity preferences and might yield a somewhat different classification than Ghirardato and Marinacci (2002), Klibanoff et al. (2005), and the EUUP model.

The ordering of events by the degree of ambiguity is determined by the second-order preference $\succsim^2$ over secondary acts in $\hat{\Delta}$. To ensure that this ordering is attributed exclusively to ambiguity (to the dispersion of the probabilities of events), the events under consideration must have the same expected probability. Otherwise, the ordering may correspond to the expected probabilities and not to ambiguity, as the DM would prefer to receive the good outcome when the event with the highest expected probability occurs. A similar approach is broadly used in comparative statics, where in order to isolate the effect of one element, all other elements are kept constant. A similar approach is also utilized to derive a measure of risk (e.g., Rothschild and Stiglitz, 1970), in which the expected outcome of all alternatives is kept constant, such that the DM‘s ordering of alternatives can be attributed solely to risk. With this notion, an ordering of events by the degree of ambiguity, induced by the second-order preference, can be defined.

Definition 2. Let the uncertain probabilities of events $E, F \in \mathcal{E}$ have the same expectation, i.e., $\mathrm{E}[\mathrm{P}(E)] = \mathrm{E}[\mathrm{P}(F)]$. F is more ambiguous than E if and only if

$$\hat{\delta}_E \succsim^2 \hat{\delta}_F$$

by any ambiguity-averse DM.

In view of $\hat{\delta}_E$ as describing the (uncertain) probability $\{\mathrm{P}(E)\}_{\mathrm{P}\in\mathcal{P}}$ of event E, $\hat{\delta}_E \succsim^2 \hat{\delta}_F$ means that any ambiguity-averse DM prefers the good outcome with the (uncertain) probability $\{\mathrm{P}(E)\}_{\mathrm{P}\in\mathcal{P}}$ to the good outcome with the (uncertain) probability $\{\mathrm{P}(F)\}_{\mathrm{P}\in\mathcal{P}}$. Since, by δ_E and δ_F, events E and F are associated with the same outcome and, by $\hat{\delta}_E$ and $\hat{\delta}_F$, they also have the same expected probability, their ordering by the second-order preference $\succsim^2$ is attributed exclusively to ambiguity. By Izhakian (2017, Proposition 1),

$$\delta_E \succsim^1 \delta_F \Longleftrightarrow \hat{\delta}_E \succsim^2 \hat{\delta}_F, \tag{7}$$

which means that $\succsim^1$ over primary indicator acts coincides with $\succsim^2$ over their associated secondary acts. Thus, an ordering of events by the degree of ambiguity as revealed by $\succsim^1$ can also be defined.

Definition 3. Let the uncertain probabilities of events $E, F \in \mathcal{E}$ have the same expectation. F is more ambiguous than E if and only if

$$\delta_E \succsim^1 \delta_F$$

by any ambiguity-averse DM.

Since δ_E and δ_F have the same possible outcomes, the assumption that $\mathrm{E}[\mathrm{P}(E)] = \mathrm{E}[\mathrm{P}(F)]$ ensures that they also have the same extent of risk and the same expected outcome, such that the only difference is the extent of ambiguity. Accordingly, their ordering by the first-order preference $\succsim^1$ is attributed solely to ambiguity. This assumption is only necessary for the extraction of a measure of ambiguity. Unlike in other models, this assumption is not required for the separation of attitudes and beliefs. Moreover, this assumption applies only to beliefs and not to attitudes toward ambiguity or risk.[30] In particular, to order events by the degree of ambiguity, risk neutrality is not assumed or implied.

[30] Other models make more restrictive assumptions on both attitudes and beliefs. For example, since in the smooth model preferences for ambiguity are outcome dependent and formed as preferences for certainty equivalent utilities,

The assumption in Definitions 2 and 3 is not unfamiliar. In risk theory, the ordering of bets (uncertain outcomes) by the degree of risk is commonly implicitly or explicitly assumed or derived to apply only to bets that have the same expected outcome. For example, Rothschild and Stiglitz (1970, Equation 2) define a bet X to be less risky than a bet Y "if X and Y have the same mean but every risk averter prefers X to Y, i.e., $\mathrm{E}[\mathrm{U}(X)] \geq \mathrm{E}[\mathrm{U}(Y)]$ for all concave U." Analogously, in Definitions 2 and 3 the ordering of two events by their uncertain probabilities is applied to events that have the same expected probability. Jewitt and Mukerji's (2017) definition of a more ambiguous event provides the same ordering as Definitions 2 and 3. Although in Klibanoff et al. (2005) an ordering of events by ambiguity is not explicitly defined, arguably, such ordering would coincide with Definition 3, since in the smooth model the utility function is normalized to $\mathrm{U}(1) = 1$.

Usually second-order preferences are not immediately observable. Definition 3 can assist the elicitation of these preferences. By Equation (7), there is a one-to-one relation between first- and second- order preferences over primary indicator acts and over their associated secondary acts. Therefore, by providing the DM with choices over primary indicator acts, the preferences over their associated secondary acts can be revealed. Thereby, second-order beliefs (probabilities), serving to form the certainty equivalent probabilities, can be elicited. Attitudes toward ambiguity, captured by the outlook function, can be elicited as well. The identification of second-order beliefs, however, is outside the scope of this paper and is left for future research. Nevertheless, when the identification of second-order beliefs is limited (or, perhaps, impossible), there is no reason not to assume that they follow a uniform distribution, as discussed below.

Definitions 2 and 3 provide an ordering of events that arises from preferences for ambiguity. These preferences, however, are derived by both beliefs about probabilities and attitudes toward their dispersions. Namely, the second-order preference $\succsim^2$ in Definition 2 reflects the DM's behavioral characteristics with respect to the probabilities of events (i.e., second-order betting behavior). The DM's personal attitude toward ambiguity, captured by the outlook function Υ, is derived from $\succsim^2$. Since $\succsim^2$ over $\hat{\Delta}$ applies exclusively to the probabilities of events independent of their associated consequences (outcome independent), by Izhakian (2017, Proposition 2),

$$\hat{\delta}_E \succsim^2 \hat{\delta}_F \Longleftrightarrow \Upsilon^{-1}\left(\int_{\mathcal{P}} \Upsilon(\mathrm{P}(E))\, d\xi\right) \geq \Upsilon^{-1}\left(\int_{\mathcal{P}} \Upsilon(\mathrm{P}(F))\, d\xi\right),$$

which implies that an ordering by the degree of ambiguity—reflecting only beliefs, not attitudes—can be defined by mean-preserving spreads in probabilities. The concept of mean-preserving spreads is applied to outcomes to define a risk ranking (Rothschild and Stiglitz, 1970), whereas here this concept is applied to probabilities to define an ambiguity ranking.

Definition 4. Event $F \in \mathcal{E}$ is more ambiguous than event $E \in \mathcal{E}$ if and only if there exists a random variable ϵ, such that

$$\mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)] =_d \mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)] + \epsilon,$$

in order to separate attitudes from beliefs, the smooth model (Klibanoff et al., 2005, p. 1861) requires preferences for risk and for certainty equivalents to remain unchanged across every subset of first-order priors. Formally, in the smooth model's notation, "(i) The restriction of $\succeq_\Pi$ to lottery acts remains the same for every support $\Pi \subset \Delta$. (ii) The same invariance with respect to Π holds for the risk preferences derived from $\succeq^2_\Pi$," where Δ is the set of priors. In contrast, in the EUUP model, the separation between attitudes (tastes) and beliefs is immediately obtained, as preferences for ambiguity are outcome independent and apply directly and solely to probabilities.

where $=_d$ means equal in distribution and $\mathrm{E}[\epsilon \mid \mathrm{P}(E)] = \mathrm{E}[\epsilon] = 0$.[31] That is, $\mathrm{P}(F)$ is a mean-preserving spread of $\mathrm{P}(E)$. If ϵ is not identically zero, then F is strictly more ambiguous than E.

Definition 4 does not assume that events share an identical expected probability or similar probability distributions. Rather, one event being more ambiguous than another is determined by the deviations of the event's possible probabilities from the respective expected probability. Sometimes Definition 4 is arguably a definition of *objective* ordering of events by ambiguity. For example: when considering a representative DM; when all DMs agree upon the possible probabilities of events and the likelihoods of these probabilities; when the DMs can convince each other about these probabilities (à la Gilboa et al., 2010); or when the DMs shape their beliefs based purely upon the information they have and avoid any aspect of subjective interpretation or judgment of the information available. Alternatively, consider a DM who adopts the provided objective probabilities as her own subjective beliefs. Suppose that the DM is informed about the set of possible priors $\mathcal{P}$ and that each $\mathrm{P} \in \mathcal{P}$ is obtained with an objective probability.[32] Plausibly, the DM would view the objective probability distributions over consequences as her subjective set of priors and would adopt the probabilities of probabilities as her second-order belief. In this respect, Halevy (2007) finds that, while facing two-stage objective lotteries, individuals may violate the reduction of compound (objective) lotteries and exhibit ambiguity-averse behavior. Segal's (1987) theory supports these findings.

The next theorem ties the ordering by preferences (attitudes) in Definitions 2 and 3 to the ordering by beliefs in Definition 4. It is guided by the notion that every ambiguity-averse DM prefers a less ambiguous event to a more ambiguous one.

Theorem 3. *Suppose that E and F are events in $\mathcal{E}$ with an identical expected probability. Then,*

$$\mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)] =_d \mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)] + \epsilon \quad \Longleftrightarrow \quad \hat{\delta}_E \succsim^2 \hat{\delta}_F \quad \Longleftrightarrow \quad \delta_E \succsim^1 \delta_F$$

by every ambiguity-averse DM, where $\mathrm{E}[\epsilon \mid \mathrm{P}(E)] = \mathrm{E}[\epsilon] = 0$. That is, Definitions 2, 3, and 4 of the more ambiguous event coincide.

Theorem 3 establishes that the probability spreads define the degree of ambiguity. Once these spreads are extracted, as suggested in Definition 4, the conditions to measure them by the variance of probabilities must be identified. Namely, the conditions under which an ambiguity-averse DM prefers a favorable outcome with a lower variance of probabilities have to be identified. Such conditions apply to the nature of either attitudes toward ambiguity or beliefs. The former describes cases of a quadratic attitude toward ambiguity or a constant absolute ambiguity aversion (CAAA), formed by the outlook function Υ.[33] The latter describes cases in which the event probabilities are either uniformly or truncated elliptically distributed, i.e., the second-order probability distribution (probabilities of probability distributions) ξ is either uniform or truncated

[31] The condition $\mathrm{E}[\epsilon \mid \mathrm{P}(E)] = \mathrm{E}[\epsilon]$ means that ϵ is mean independent of $\mathrm{P}(E)$. Note that equality in distribution is a much weaker condition than equality, and that mean independence is weaker than independence.

[32] For instance, this can be generated by a randomizing device, such as a roulette wheel, that determines the number of balls in an Ellsberg urn.

[33] CAAA implies that the ambiguity attitude is not sensitive to a positive probability scaling, and is formed by the outlook function $\Upsilon(\mathrm{P}(E)) = \frac{1-e^{-\eta \mathrm{P}(E)}}{\eta}$ (Izhakian, 2017).

elliptical.[34] The probability $\mathrm{P}(E)$ of event E is elliptically distributed if its probability characteristic function has the following form

$$\phi_{\mathrm{P}(E)}(t) = e^{it\mathrm{E}[\mathrm{P}(E)]}\Psi\left(\frac{1}{2}t^2\mathrm{Var}[\mathrm{P}(E)]\right),$$

where $i=\sqrt{-1}$ and Ψ is a characteristic generator.

Theorem 4. *Suppose that E and F are events in $\mathcal{E}$ with an identical expected probability. Then,*

$$F \text{ is more ambiguous than } E \quad \Longleftrightarrow \quad \mathrm{Var}[\mathrm{P}(F)] \geq \mathrm{Var}[\mathrm{P}(E)],$$

when one or more of the following conditions hold true:

(i) *The probabilities of E and F are uniformly distributed;*
(ii) *The probabilities of E and F are truncated elliptically distributed with an identical characteristic generator;*
(iii) *The DM's attitude toward ambiguity is of the CAAA type and $\mathrm{P}(E)$ is close to $\mathrm{E}[\mathrm{P}(E)]$*[35]*;*
(iv) *The DM's attitude toward ambiguity is quadratic.*

Theorem 4 identifies the conditions under which the ordering of events by the variance of their probabilities coincides with the ordering by ambiguity of Definitions 2, 3, and 4. Condition (i) in Theorem 4 is that the probabilities of events are uniformly distributed (i.e., a uniform second-order belief). When the identification of second-order beliefs is limited (or, perhaps, impossible), there is no reason not to assume that these beliefs follow a uniform distribution. This is equivalent to assuming that the DM does not have superior information to infer, or reason to assume a greater likelihood of a particular probability distribution, and therefore assigns equal likelihoods to each possible distribution. Consider, for example, an urn with 10 colored balls that are either black or yellow, in an unknown proportion; there is no reason for the DM to believe that the 4-6 proportion of balls is more likely than the proportion 5-5. An assumption of a uniform second-order belief is consistent with the principle of insufficient reason, which asserts that given n possibilities that are indistinguishable except for their names, each possibility should be assigned a probability equal to $\frac{1}{n}$ (Bernoulli, 1713; Laplace, 1814). It is also consistent with the idea of the simplest non-informative prior in Bayesian probability (Bayes et al., 1763), where equal probabilities are assigned to all possibilities; and the principle of maximum entropy (Jaynes, 1957), which asserts that the probability distribution that best describes the current state of knowledge is the one with the largest entropy.

Further, the alternative condition in Theorem 4, that the probabilities of events are truncated elliptically distributed, is by no means too restrictive since the family of elliptical distribution clusters a wide range of well-known classes of distributions including, inter alia, the normal distribution, the Student-t distribution, the logistic distribution, the exponential power distribution,

[34] Truncated distributions are considered because probability values are bounded. A truncated elliptical distribution takes the form $\varphi(x;\mu,\sigma,a,b) = \frac{\frac{c}{\sigma}\psi\left(\frac{x-\mu}{\sigma}\right)}{\int_{-\infty}^{b}\frac{c}{\sigma}\psi\left(\frac{x-\mu}{\sigma}\right)dx-\int_{-\infty}^{a}\frac{c}{\sigma}\psi\left(\frac{x-\mu}{\sigma}\right)dx}$, where $[a,b]\in\mathbb{R}$ is the support of φ; ψ is a density generator; and c is a normalizing constant.

[35] "Close" in the sense that the third and higher absolute central moments of $\mathrm{P}(E)$ are of a smaller order than the second absolute central moment, and are therefore negligible.

and the Laplace distribution. Moreover, this assumption falls far short of the common assumption of normally distributed outcomes used to measure risk. Most theoretical and empirical studies of risk and its implications are (explicitly or implicitly) underpinned by the assumption of normally distributed outcomes; for instance, the capital asset pricing model (CAPM, Sharpe, 1964; Lintner, 1965; Mossin, 1966), the intertemporal CAPM (Merton, 1973), the option pricing model (Black and Scholes, 1973), and its expansion (Merton, 1976).

A similar argument supports the alternative condition in Theorem 4 of a quadratic or a CAAA outlook function, Υ. To measure risk using the variance of outcomes, a quadratic utility function is assumed (e.g., Tobin, 1958; Markowitz, 1959; Rothschild and Stiglitz, 1970). Otherwise, given two lotteries with the same expected outcome, a risk-averse DM may prefer the lottery with the higher variance. This assumption underpins the widely used mean-variance analysis and modern portfolio theory.

5. Ambiguity measurement

The well-defined ordering of events by their degree of ambiguity, measured by $\mathrm{Var}\left[\mathrm{P}\left(\cdot\right)\right]$, facilitates an ordering of primary acts, which is a necessary step toward extracting a measure of ambiguity. To define such an ordering, ambiguous primary acts have to first be identified.

Definition 5. A primary act $f \in \mathcal{F}_0$ is unambiguous if and only if for any $x \in X$ the event $\{s \in \mathcal{S} \mid f(s) = x\}$ is unambiguous. That is, $\{s \in \mathcal{S} \mid f(s) = x\} \in \mathcal{H}$, for every $x \in X$.

Definition 5 implies that a primary act is unambiguous if and only if it is a roulette lottery (i.e., the probability of each of its outcomes is known and objective). The first-order preference relation $\succsim^1$ over unambiguous primary acts coincides with the definition of the "unambiguous preference" relation in Ghirardato et al. (2004, Definition 3), and may be used to endogenously identify the set of priors $\mathcal{P}$. Furthermore, the second-order preference relation $\succsim^2$ may be represented as an extension of this unambiguous preference.[36] The endogenous identification of the set of priors, however, is outside the scope of this paper and is left for future research.

Using a comparative statics approach, the ordering of primary acts considers acts that have the same expected probabilities of outcomes (implying the same support, the same expected outcome, and the same risk), such that the only difference between the acts is the dispersion of probabilities around their expectations. This approach attributes the ordering, as revealed from the DM's choices, exclusively to ambiguity and related preferences. Otherwise, the ordering can be attributed to the risk level or the level of the expected outcome.

Definition 6. Let $f, g \in \mathcal{F}_0$ be primary acts under which the expected probabilities of each consequence $x \in X$ are identical, i.e., $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_g(x)\right]$ for every $x \in X$. Act g is more ambiguous than f if and only if

$$f \succsim^1 g,$$

by any ambiguity-averse DM.

Definition 6 extends Definition 3 from indicator acts to primary acts in $\mathcal{F}_0$. As in Definition 3, the assumption that $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_g(x)\right]$ for every $x \in X$ ensures that the ordering of acts

[36] The author thanks an anonymous referee for suggesting this insight.

can be attributed exclusively to ambiguity, since the comparison is between acts that have the same extent of risk, the same expected outcome, and differ only in the extent of ambiguity. As in Definitions 2 and 3, this assumption is analogous to the assumption in Rothschild and Stiglitz's (1970) theoretical foundation of risk measurement; the ordering of bets (uncertain outcomes) by the degree of risk (and, in some cases, by the volatility of outcomes) considers only bets that have the same expected outcome. Also, as in Definitions 2 and 3, the assumption in Definition 6 applies only to beliefs and not to attitudes toward ambiguity or risk. In particular, to order primary acts by their degree of ambiguity, it is not assumed or implied that the DM is risk-neutral, only that the relevant acts bear the same risk.

In Definition 6, the partition of the state space $\mathcal{S}$ into events, induced by each primary act, may be different. Namely, for a given $x \in X$, $f^{-1}(x)$ is not necessarily equal to $g^{-1}(x)$. Definition 6 is permitted since the consistency of first- and second-order preferences in the EUUP model applies only to primary indicator acts and their associated secondary acts. This consistency is less restrictive in the EUUP model than in other models, as it is not enforced over all primary acts in $\mathcal{F}_0$ and secondary acts in $\widehat{\mathcal{F}}$. For example, in the smooth model (Klibanoff et al., 2005, Assumption 3), the consistency axiom enforces $f \succsim^1 g \iff f^2 \succsim^2 g^2$ over all acts, where f^2 identifies the conditional certainty equivalents of f. Thus, under the smooth model, even when $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_g(x)\right]$ for every $x \in X$, the ordering $f^2 \succsim^2 g^2$ cannot be attributed solely to ambiguity, since f^2 and g^2 are subject to outcomes, risk, and risk preferences. Namely, in the smooth model, the second-order preference $\succsim^2$ is outcome dependent, as elaborated in Appendix A.1.

Jewitt and Mukerji (2017) suggest that act g is more ambiguous than act f if g disadvantages an ambiguity-averse DM but does not affect an ambiguity-neutral DM. Their definition generates the same ordering of acts as Definition 6. Loosely speaking, a DM who is risk-averse, ambiguity-neutral, and g is not disadvantageous to her, implies that f and g have the same expected outcome and the same risk, which in turn implies the same expected outcome probabilities. In contrast, a DM who is ambiguity-averse and g is disadvantageous to her, implies that due to the ambiguity and her aversion to it, under g the perceived probabilities of favorable outcomes are lower and the perceived probabilities of unfavorable outcomes are higher, resulting in a lower expected utility.

A valid measure of ambiguity must provide an ordering of primary acts that coincides with the ordering provided by an ambiguity-averse DM. Satisfying this requirement, the next theorem introduces a new measure of ambiguity. This theorem, which is the central result of the current paper, asserts that the degree of ambiguity associated with a primary act can be measured by the expected volatility of the act's related probabilities.

Theorem 5. *Suppose that the conditions of Theorems 2 and 4 are satisfied, and an ambiguity-averse DM whose preferences are reference independent. Then,*

$$f \succsim^1 g \quad \iff \quad \mho^2[f] \leq \mho^2[g],$$

for every act $f, g \in \mathcal{F}_0$ under which the expected probabilities of each consequence $x \in X$ are identical, and where

$$\mho^2[f] \equiv \int_X \mathrm{E}\left[\varphi_f(x)\right] \mathrm{Var}\left[\varphi_f(x)\right] dx. \tag{8}$$

Theorem 5 links the measure of ambiguity, denoted $\mho^2$ (mho^2), to the DM's preferences, formed by the dual representation expanded in Theorem 2. The concept that ambiguity takes the form of probability perturbations, and aversion to ambiguity takes the form of aversion to mean-preserving spreads in probabilities, underpins $\mho^2$. Theorem 5 proves that, given two acts, identical except for their degree of ambiguity, any ambiguity-averse DM with reference-independent preferences would prefer the act with the lower $\mho^2$ to the act with the higher $\mho^2$. That is, she would prefer the act whose associated probabilities are on average less volatile (lower spread). $\mho^2$ aggregates the probability variances, which measure the dispersions of the probabilities of each outcome, while assigning a weight equal to the probability's expectation to the variance of each outcome's probability. Therefore, $\mho^2$ can be referred to as the *expected volatility of probabilities* and written as

$$\mho^2[f] \equiv \mathbb{E}\left[\operatorname{Var}\left[\varphi_f(x)\right]\right],$$

where the expectation $\mathbb{E}[\cdot]$ is taken using the expected probabilities.

The accuracy of $\mho^2$ is derived from the accuracy of the expansion in Theorem 1, which is of order $o\left(\mathrm{E}\left[|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]\right)$ as $|\varphi(x)-\mathrm{E}[\varphi(x)]| \to 0$. Importantly, this expansion is one order of magnitude (relative to the expanded variable) more accurate than the quadratic expansion used in risk measurement, $o\left(\mathbb{E}\left[|x-\mathbb{E}[x]|^2\right]\right)$ (e.g., Pratt, 1964; Arrow, 1965). Theorem 5 employs the assumptions of Theorem 4, as discussed in Section 4; in particular, the assumption that the probabilities of outcomes are uniformly or truncated elliptically distributed with the same characteristic generator. As with respect to ordering events by the volatility of probabilities, the assumption of uniformly distributed probabilities is equivalent to the assumption that the DM does not have superior information to infer, or reason to assume, a greater likelihood of a particular probability distribution, and thus assigns equal likelihoods to each possible distribution. As discussed in Section 4, this assumption is consistent with the principle of insufficient reason (Bernoulli, 1713; Laplace, 1814); the idea of the simplest non-informative prior in Bayesian probability (Bayes et al., 1763); and the principle of maximum entropy (Jaynes, 1957). The alternative assumption of truncated elliptically distributed probabilities is significantly less restrictive than the assumption of normally distributed outcomes, which is required to measure risk, since among many other classes of distributions, the family of elliptical distributions also includes the class of normal distributions.

Theorem 5 assumes reference-independent preferences, meaning CEU preferences (Schmeidler, 1989). This assumption implies that the DM's reference point is either $k=\min_{x\in X} x$ or $k=\max_{x\in X} x$, i.e., either all outcomes are considered favorable or all are considered unfavorable. The former can be described as if the minimum utility is 0; the latter, as if the maximum utility is 0. An assumption about beliefs can replace the assumption in Theorem 5 about reference independence; in particular, by the assumption that outcomes are symmetrically distributed. Such a symmetry in a framework with uncertain probabilities can be defined as follows.

Definition 7. The outcomes of an act $f \in \mathcal{F}_0$ are symmetrically distributed around a point of symmetry k if and only if, for any $x, y \in X$ that satisfy $|x-k| = |y-k|$,

$$\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_f(y)\right] \quad \text{and}$$
$$\mathrm{E}\left[\left(\varphi_f(x)-\mathrm{E}\left[\varphi_f(x)\right]\right)^n\right] = \mathrm{E}\left[\left(\varphi_f(y)-\mathrm{E}\left[\varphi_f(y)\right]\right)^n\right],$$

for $n = 2, 3, 4, \ldots$.[37]

With this Definition 7 in place, Theorem 5 can be restated. Note that in the next theorem, the more restrictive assumption of normally distributed outcomes, which is widely used to measure risk using the variance of outcomes, is not required. To measure ambiguity, an assumption of symmetrically distributed outcomes is sufficient.

Theorem 6. *Suppose that the conditions of Theorem 2 and 4 are satisfied, and an ambiguity-averse DM. Then,*

$$f \succsim^1 g \quad \Longleftrightarrow \quad \mho^2 [f] \le \mho^2 [g],$$

for every act $f, g \in \mathcal{F}_0$ whose outcomes are symmetrically distributed around a reference point k, and under which the expected probabilities of each consequence $x \in X$ are identical.

In the case of acts taking finitely many values, $\mho^2$ is formed by[38]

$$\mho^2 [f] \equiv \sum_j \mathrm{E}\left[\varphi_f\left(x_j\right)\right] \mathrm{Var}\left[\varphi_f\left(x_j\right)\right]. \tag{9}$$

This measure can also be applied to a nonempty subset $Y \subset X \subseteq \mathbb{R}$ of consequences,

$$\mho^2 [f, Y] \equiv \int_Y \mathrm{E}\left[\varphi_f(y)\right] \mathrm{Var}\left[\varphi_f(y)\right] dy,$$

or to a given event $E \in \mathcal{E}$,

$$\mho^2 [f, E] \equiv \int_{f^{-1}(x)\in E} \mathrm{E}\left[\varphi_f(x)\right] \mathrm{Var}\left[\varphi_f(x)\right] dx.$$

6. Properties of $\mho^2$

6.1. The extent of ambiguity

It is worthwhile to demonstrate some properties of $\mho^2$. Consider a *large* urn with 30 colored balls, which are either black or yellow, in an unknown proportion. Assume that drawing a black ball (B) entitles the DM to a sum of \$0, and a yellow ball (Y) entitles her to a sum of \$1. The probability of B can be one of the values $\frac{0}{30}, \frac{1}{30}, \ldots, \frac{30}{30}$, where the DM, who has no information indicating which of the possible probability values is more likely, acts as if each is equally likely (i.e., the principle of insufficient reason). Since every $\mathrm{P} \in \mathcal{P}$ is additive, the probability of Y can also be one of the values $\frac{0}{30}, \frac{1}{30}, \ldots, \frac{30}{30}$, with equal likelihoods. Thus, by Equation (9), the degree of ambiguity (in units of probability) is $\mho = 0.298$. Now, consider a smaller urn with only 10 colored balls, which are either black or yellow, in an unknown proportion. In this case, the probability of each event (B or Y) can be one of the values $\frac{0}{10}, \frac{1}{10}, \ldots, \frac{10}{10}$. Thus, the ambiguity

[37] Since the measure $\mho^2$ considers only the first two moments of the distribution of probabilities, $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_f(y)\right]$ and $\mathrm{Var}\left[\varphi_f(x)\right] = \mathrm{Var}\left[\varphi_f(y)\right]$ are sufficient.

[38] Where $\varphi_f\left(x_j\right)$ denotes the (uncertain) probability mass function.

Table 1
Degrees of ambiguity.

#Balls						
Total	Y	B	P	Q(Y)	V	$\mho$
30	15	15	$\frac{15}{30}$	0.500	0.316	0.000
∞	$0, 1, \ldots$	$0, 1, \ldots$	$0 \ldots 1$	0.445	0.281	0.289
30	$0, 1, \ldots, 30$	$0, 1, \ldots, 30$	$\frac{0}{30}, \frac{1}{30}, \ldots, \frac{30}{30}$	0.435	0.275	0.298
10	$0, 1, \ldots, 10$	$0, 1, \ldots, 10$	$\frac{0}{10}, \frac{1}{10}, \ldots, \frac{10}{10}$	0.417	0.263	0.316
1	$0, 1$	$0, 1$	$0, 1$	0.250	0.158	0.500

associated with a bet on this urn, $\mho = 0.316$, is higher than for the larger urn. If there is only one ball in the urn, of unknown color, then $\mho = 0.5$. In the other extreme case, if there is an infinite number of balls in the urn, then $\mho = 0.289$. Table 1 is a stylized description of these variations. In this table, the perceived probabilities (Q) and the utility values (V) are computed, respectively, by Equations (3) and (2), assuming a constant relative ambiguity aversion $\Upsilon(\mathrm{P}(E)) = \sqrt{\mathrm{P}(E)}$, a constant absolute risk aversion $\mathrm{U}(x) = 1 - e^{-x}$, and a reference point $k = 0$.[39]

Observe that a larger number of possible probability values (of each event), which in this case are uniformly spread over the interval [0, 1], implies a lower degree of ambiguity. To illustrate the intuitiveness, note that since the favorable (unfavorable) outcome is identical for all bets, a choice of urns is actually a bet on the composition of the urn rather than on the outcome. Suppose that the DM chooses to bet on the 10-ball urn and her probability (proportion of balls) assessment is wrong. The minimum magnitude of her error (in terms of probability) is $\frac{1}{10}$. If, however, she chooses to bet on the 30-ball urn and her probability assessment is wrong, the minimum magnitude of her error is only $\frac{1}{30}$. The maximum magnitude of her error with respect to both urns is 1. The average size of the error, which accounts for the magnitudes and the likelihoods of all errors, is also greater for the 10-ball urn than for the 30-ball urn. For example, suppose that the DM wrongly assesses the probability of gain by 50%. The average magnitude of the error concerning the 10-ball urn is $2\sum_{i=1}^{5} \frac{i}{10}\frac{1}{11} = \frac{3}{11}$, while for the 30-ball urn it is only $2\sum_{i=1}^{15} \frac{i}{30}\frac{1}{31} = \frac{8}{31}$. Similarly, the degree of ambiguity of the 30-ball urn is lower than that of the 10-ball urn. The next observation formally defines the property that arises from this intuition.

Observation 1. *Suppose a non-empty, finite set of probability measures $\mathcal{P}$, such that the marginal probabilities $\varphi_f(x)$ of each $x \in X$ are uniformly distributed over some interval $[a_x, b_x]$. Then, the higher the cardinality of $\mathcal{P}$, the lower the degree of ambiguity $\mho^2[f]$.*

The next observation identifies the range of values that $\mho^2$ can obtain concerning primary acts taking finitely many values.

Observation 2. *Suppose a primary act f that takes finitely many values. The act's degree of ambiguity, $\mho^2[f]$, then ranges between 0 and $\frac{1}{4}$.*

The minimum degree of ambiguity, $\mho^2[f] = 0$, is attained if and only if act f is unambiguous, i.e., $\{s \in \mathcal{S} \mid f(s) = x\} \in \mathcal{H}$ for every $x \in X$. Concerning acts that take finitely many values,

[39] Recall that the utility function is normalized such that $\mathrm{U}(k) = 0$.

the maximum degree of ambiguity is attained when there are only two possible outcomes and the probability of each is either 0 or 1 with equal likelihoods. In this case, by Equation (9), the ambiguity attains the value $\mho^2 = \frac{1}{4}$. The next observation analyzes the values of $\mho^2$ when primary acts take infinitely many values.

Observation 3. *Suppose a primary act f that takes infinitely many values. The act's degree of ambiguity, $\mho^2 [f]$, then ranges between 0 and ∞.*

6.2. Outcome and risk independence

One of the most important properties of $\mho^2$ is outcome independence (up to a state space partition) and, thereby, risk independence. Given an event, the degree of ambiguity is invariant to the consequence of this event. That is, changing the consequence associated with an event does not affect its degree of ambiguity. Consider, for example, an event with an unknown probability of a \$100 payoff. Changing the magnitude of the payoff to \$1,000 should not affect the event's perceived probability or the degree of ambiguity, since no new information about likelihoods has been obtained. This property of outcome independence holds true not only for single events, but also for primary acts.

Observation 4. *Suppose that acts f and g induce the same partition of the state space $\mathcal{S}$ into events. Then, given a fixed set of probability measures $\mathcal{P}$ on $\mathcal{S}$, $\mho^2 [f] = \mho^2 [g]$.*

This observation immediately implies the following.

Observation 5. *Given a fixed set of probability measures $\mathcal{P}$ on $\mathcal{S}$, $\mho^2$ is invariant to the outcomes associated with events (outcome independence) up to a state space partition.*

That is, when modifying the outcomes assigned by an act, as long as both the act's induced partition of the state space into events and the set of probability measures on the state space remain unchanged, the degree of ambiguity remains unchanged (but the degree of risk might change). Consider, for example, a linear shift of the outcomes assigned by act f, such that $f^* = \alpha + f$. In this case, $\mho^2 \left[f^*\right] = \mho^2 [f]$. This property of $\mho^2$ holds for any strictly monotonic transformation of outcomes, as long as the set of probability measures is invariant to the changes in outcomes.

Observation 6. *Let act f^* be a strictly (increasing or decreasing) monotonic transformation of act f. Given a fixed set of probability measures $\mathcal{P}$ on $\mathcal{S}$, $\mho^2 \left[f^*\right] = \mho^2 [f]$.*

The intuition for this property is that, conceptually, $\mho^2$ aggregates the ambiguity of events (assigned with different weights) independently of their associated outcomes. Since f and its strictly monotonic transformation f^* induce the same partition of the state space $\mathcal{S}$ into events, they have the same degree of ambiguity; however, they might have a different degree of risk. Assume, arguendo, that ambiguity should be affected by the magnitude of outcomes. For example, given a two-color Ellsberg urn, different payoffs assigned to each color should result in a different degree of ambiguity. Suppose that the payoff assigned to each event (color) is doubled. Risk (measured by the standard deviation of outcomes) is then doubled and, accordingly, ambiguity should also increase. But if so, ambiguity is not risk independent. By $\mho^2$, different acts may have different extents of ambiguity only if they induce different partitions over the state space. In a two-color Ellsberg urn, different partitions of the state space may be induced only if one act assigns the same payoff to both colors, and the other assigns a different payoff to each color.

Risk independence is a critical property, because it allows for the detection of the distinct impact of ambiguity in isolation of risk. Other measures that were taken as proxy for ambiguity (e.g., the variance of the mean or the variance of the variance) are outcome- and risk dependent, as discussed in Subsection 6.5. Notably, decision-making processes consider not only the extent of ambiguity, but also the extent of risk. Hence, when making choices, these two elements share a common role. For example, suppose that $f \succsim^1 g$, and let f^* induce the same partition of the state space as f. Then, the ambiguity of f^* is the same as the ambiguity of f, but the risk of f^* may be higher, such that $f^* \precsim^1 g$.

6.3. Ambiguity hedging

An interesting property of $\mho^2$ is that ambiguity can be hedged (canceled out). This can happen when composing a "portfolio" of acts. To demonstrate, consider acts $f = \left(E : 0, E^C : x\right)$ and $g = \left(E^C : 0, E : x\right)$. Even if separately each act has a positive degree of ambiguity, a portfolio consisting of only these two acts has zero ambiguity (and in this case, also zero risk). The reason is that $f + g$ induces a different partition of the state space $\mathcal{S}$ than each act separately. In particular, $(f + g)^{-1}(x) = E \cup E^C = \mathcal{S}$, while $f^{-1}(x) = E^C$ and $g^{-1}(x) = E$. Since $\mathrm{P}(\mathcal{S})$ is always one, the degree of ambiguity of the entire state space, measured by $\mathrm{Var}[\mathrm{P}(\mathcal{S})]$, is zero.

Consider acts $f = \left(E^C : 0, E : x\right)$ and $g = \left(F^C : 0, F : x\right)$, where E and F are mutually exclusive. In this case, the ambiguity associated with x may be lower under $f + g$ than under f or g separately. This happens when the probabilities of E and F are negatively correlated ($\mathrm{Cov}[\mathrm{P}(E), \mathrm{P}(F)] < 0$), which implies that unpacking $E \cup F$ into separate events E and F (associated with different outcomes) increases the ambiguity.[40] Note that, since the probability of any event is perfectly negatively correlated with the probability of its complementary event (Lemma 1), the ambiguity of a union of an event and its complementary event is always zero.

Ambiguity might be hedged as a result of changes in the partition of the state space into events by packing or unpacking events. This can be demonstrated by the Ellsberg three-color experiment. In this experiment, the DM is presented with an urn that contains 90 colored balls, 30 of them red and the others either black or yellow in an unknown proportion. A ball will be drawn from the urn at random and the payoff for a correct bet is \$100. The experiment consists of two parts. First, the DM has to choose between two bets: the next drawn ball is red (R), or the next drawn ball is black (B), formed respectively by acts f and g. Then she has to choose between betting that the next drawn ball is red or yellow (RY) or, alternatively, that the next drawn ball is black or yellow (BY), formed respectively by acts f^* and g^*. Suppose that the DM does not have any information indicating which of the possible urn compositions (probabilities) is more likely, and thus acts as if she assigns an equal likelihood to each possibility (i.e., the principle of insufficient reason). Table 2 describes the ambiguity associated with each event and each act. The ambiguity of the event with the high payoff under each act is underlined.

The findings of behavioral experiments show that individuals typically prefer R to B and BY to RY; formally, $f \succsim^1 g$ and $g^* \succsim^1 f^*$. Table 2 illustrates that, in accordance with Theorem 5,

[40] *Support theory* (Tversky and Koehler, 1994; Rottenstreich and Tversky, 1997) documents that the judged probability of an event generally increases when its description is unpacked into disjoint components.

Table 2
Ellsberg's three-color experiment.

Act	Prize ($)			Event $\mho$					Act $\mho$
	R	Y	B	R	Y	B	RY	BY	
f	100	0	0	$\underline{0.000}$	0.113	0.113			0.000
g	0	0	100	0.000	0.113	$\underline{0.113}$			0.196
f^*	100	100	0				$\underline{0.159}$	0.000	0.196
g^*	0	100	100				0.159	$\underline{0.000}$	0.000

$\mho[f] < \mho[g]$ and $\mho\left[g^*\right] < \mho\left[f^*\right]$. Namely, DMs usually prefer a less ambiguous bet, which implies an ambiguity-averse behavior. Table 2 also illustrates that under act g, event BY can be described as unpacked into events B and Y (with different outcomes), such that the ambiguity of g is greater than that of f.[41] Under act f^*, event BY can also be described as unpacked into events B and Y, such that the ambiguity of f^* is greater than that of g^*. Formally, define three acts: $f = (Y \cup B : 0, R : 100)$, $g = (Y \cup R : 0, B : 100)$, and $h = (R \cup B : 0, Y : 100)$. Then, it can be written $f^* = f + h$ and $g^* = g + h$. In this representation, $g + h$ perfectly hedges ambiguity.

The idea that ambiguity can be hedged (diversified away) coincides with the notion of *complementarity*, which suggests that a change of the outcome assigned (by an act) to some event "causes a reversal of rankings, 'then' this reversal is due to complementarities between events (...) rather than between outcomes" (Epstein and Zhang, 2001, p. 275). In different settings, Ghirardato et al. (2004) show that the DM combines acts to optimally hedge against ambiguity. Siniscalchi (2009) defines conditions for perfect hedging against ambiguity. A similar concept is also employed by Gilboa and Schmeidler (1989) and Gilboa et al. (2010), who use a mixture of acts that hedges against uncertainty (risk and ambiguity) to demonstrate that the independence axiom is not as normatively appealing for subjective rationality as for objective rationality. In fact, the idea of ambiguity hedging underpins their uncertainty aversion axiom. Izhakian and Zender (2018) apply the concept of ambiguity hedging to the principal-agent problem, and show that the efficiency of a contract can be improved by designing a contract that induces a particular partition of the state space that minimizes the ambiguity faced by the agent (or by the principal).

6.4. *Subjective versus objective ambiguity*

The $\mho^2$ measure quantifies the ambiguity associated with *beliefs* independently of subjective attitudes toward ambiguity and risk, and independently of subjective perception of unfavorable and favorable outcomes. Beliefs are typically subjective. Even when probabilities are given, individuals may not always accept these objective probabilities as their own beliefs, and make decisions based on their subjective perception (or distortion) of probabilities (Kahneman and Tversky, 1979). In reality, most decisions are mainly based on subjective judgment of likelihoods and, respectively, subjective degrees of ambiguity. In this respect, $\mho^2$ is a measurement tool that can be used to evaluate the subjective extent of ambiguity, upon which each individual bases decisions.

Although $\mho^2$ can be viewed as a measure of subjective ambiguity, sometimes this measure can be interpreted as a measure of *objective* ambiguity. An interpretation of $\mho^2$ as a measure of

[41] Note that the probabilities of B and Y are perfectly negatively correlated.

objective ambiguity is valid when Definition 4 is interpreted as a definition of objective ordering by the degree of ambiguity. In particular, suppose that the DM, who is informed about the set of possible probability measures and their likelihoods, views these objective probabilities as her own subjective beliefs. In this case, $\mho^2$ can be viewed as measuring objective ambiguity over two-stage (objective) lotteries. Recall that Halevy (2007) finds that individuals are likely to violate the reduction of compound (objective) lotteries and exhibit ambiguity-averse behavior. $\mho^2$ can also be interpreted as a measure of objective ambiguity when beliefs are based purely on the information available, and do not involve any subjective interpretation or judgment. In this case, each DM may have her own set of priors equipped with likelihoods, which reflect only the available information. Note that higher quality information does not necessarily imply a "smaller" set of priors or a "narrower" second-order distribution, which would result in a lower $\mho^2$, but rather reflects a clearer picture of reality (Epstein and Schneider, 2007). Thus, each DM may have a different $\mho^2$, which reflects her exposure to a particular level of objective ambiguity. $\mho^2$ can also be interpreted as a measure of objective ambiguity when all DMs have the same information shaping their beliefs.

Another case where $\mho^2$ can be viewed as a measure of objective ambiguity is when all DMs agree upon the possible probabilities of events and the likelihoods of these probabilities. The DMs might initially agree upon these probabilities, or agree after discussing with each other. In the latter case, the DMs are assumed to be objectively rational à la Gilboa et al. (2010).[42] $\mho^2$ can also be viewed as a measure of objective ambiguity in representative-agent models. A representative agent can be defined, for example, as one whose attitudes and beliefs are such that if all investors in the economy had identical attitudes and beliefs, the equilibrium in the economy would remain unchanged (à la Constantinides, 1982). Such an agent faces ambiguity, which can be measured by $\mho^2$ and referred to as objective ambiguity.

6.5. Alternative measures

Since the seminal works of Knight (1921) and Ellsberg (1961), several attempts have been made to define a measure of ambiguity. Hansen et al. (1999) and Hansen and Sargent (2001), for example, interpret relative entropy as a measure of ambiguity (or of model uncertainty).[43] Similar to $\mho^2$, relative entropy is also an outcome- and risk-independent measure. It depends only on the probabilities of events, not their associated outcomes. However, while relative entropy is a measure of the distance of a single prior relative to a reference distribution, $\mho^2$ can be applied to multiple priors when a reference distribution cannot be identified or does not exist. In addition, while relative entropy is not symmetric (i.e., the entropy of P relative to Q may not be equal to the entropy of Q relative to P), $\mho^2$ is symmetric.

Some studies use the variance of a single moment of the probability distribution as a measure of ambiguity. In particular, they use the variance of the mean (e.g., Maccheroni et al., 2013) or the variance of the variance (e.g., Faria and Correia-da Silva, 2014) as measures of ambiguity. $\mho^2$ is broader than either of these measures as it accounts for both, as well as for the variances

[42] Namely, the DMs are able to convince others that their perception of the set of possible priors and the likelihoods of these priors are correct and, accordingly, so are the choices they make. A DM in Gilboa et al.'s (2010) framework is assumed to be persuadable if she is embarrassed to state her preferences.

[43] Formally, the entropy of distribution P relative to distribution Q is defined by $D_{KL}(\mathrm{P}\,|\,\mathrm{Q}) \equiv \int p(x) \ln \frac{p(x)}{q(x)} dx$, where p and q are, respectively, the densities of P and Q.

of all higher moments of the probability distributions (i.e., skewness, kurtosis, etc.), through the variance of probabilities. In this respect, $\mho^2$ is in line with Epstein and Ji (2013, 2014), who show that the variance of the variance (ambiguous variance) is an important element in decision-making processes, and should not be disregarded.

Critically, the variance of the mean and the variance of the variance are both outcome dependent and are, therefore, risk dependent. As such, neither allows for measuring ambiguity in isolation of risk. To illustrate, consider a bet that pays \$100 in an event with an unknown probability of occurrence. Presumably, although changing the payoff to \$1,000 increases the risk, the ambiguity should remain unaffected, since no new information about likelihoods has been obtained. However, both the variance of the mean and the variance of the variance indicate that the \$1,000 bet is more ambiguous than the \$100 bet, even though both bets are on the same event and the change of its associated outcome from \$100 to \$1,000 does not affect its probabilities. In contrast, $\mho^2$ indicates that both bets have the same degree of ambiguity, since $\mho^2$ is solely a function of probabilities.

Measuring ambiguity by the exclusive use of either the variance of the mean or the variance of the variance fails when the set of priors consists of distributions with either an identical mean or an identical variance. The comparison of two bets with different degrees of ambiguity and an equal mean or an equal variance is not possible when using either of these two measures. In contrast, $\mho^2$ allows for the comparison of two bets with different degrees of ambiguity and an identical mean or an identical variance, since it accounts for both an ambiguous mean and an ambiguous variance.

To illustrate, consider two bets with the same mean: bet A with the outcomes $x = (-1, 0, 1)$ and, respectively, the probabilities $\mathrm{P} = (0.5, 0, 0.5)$, and bet B with the same outcomes, but with two equally likely probability distributions, $\mathrm{P}_1 = (0.4, 0.2, 0.4)$ and $\mathrm{P}_2 = (0.3, 0.4, 0.3)$. The expected outcome of A is $\mathbb{E}_{\mathrm{P}}[x] = 0$; the expected outcome of B is either $\mathbb{E}_{\mathrm{P}_1}[x] = 0$ or $\mathbb{E}_{\mathrm{P}_2}[x] = 0$, respectively, contingent upon P_1 and P_2. The variance of the mean indicates that both A and B have a zero degree of ambiguity (are unambiguous). However, by definition (and as $\mho^2$ indicates), B, which has a positive degree of ambiguity, is more ambiguous than A, which is clearly unambiguous.

The use of the variance of the variance as a measure of ambiguity creates a similar issue. To illustrate this, consider the following: bet A with the outcomes $x = (-1, 0, 1)$ and, respectively, the probabilities $\mathrm{P} = (0.48, 0.04, 0.48)$, and bet B with the same outcomes but with two equally likely probability distributions, $\mathrm{P}_1 = (0.6, 0, 0.4)$ and $\mathrm{P}_2 = (0.4, 0, 0.6)$. The variance of the outcomes of A is $\mathbb{V}\mathrm{ar}_{\mathrm{P}}[x] = 0.96$; the variance of the outcomes of B is either $\mathbb{V}\mathrm{ar}_{\mathrm{P}_1}[x] = 0.96$ or $\mathbb{V}\mathrm{ar}_{\mathrm{P}_2}[x] = 0.96$, respectively contingent upon P_1 and P_2. The variance of the variance indicates that both A and B have a zero degree of ambiguity. However, by definition (and as $\mho^2$ indicates), B is more ambiguous than A.

Other approaches that have been proposed to estimate the degree of ambiguity confound it with attitudes toward ambiguity and are, therefore, difficult to implement using the data. For example, the sum of the nonadditive probabilities (Dow and Werlang, 1992); the discrimination among different levels of likelihood (Baillon et al., 2018); and the matching nonadditive probabilities indices (Baillon and Bleichrodt, 2015).

7. From theory to practice

The measure of ambiguity, $\mho^2$, is a practical instrument in both theoretical and empirical studies of ambiguity. This section demonstrates the applicability of $\mho^2$ in theoretical studies by utilizing it in asset pricing. Then, it demonstrates the applicability of $\mho^2$ in empirical studies, by providing a description of an empirical implementation of $\mho^2$ using the data.

7.1. Application for asset pricing

The price a DM is willing to pay for an asset could be affected by the uncertainty about the probabilities of future payoffs. She might require a premium for bearing ambiguity, in addition to the premium for bearing risk. The *risk premium* can be described as the premium that a DM is willing to pay to exchange a risky asset for a riskless one with an identical expected payoff. Similarly, the *ambiguity premium* can be described as the premium that she is willing to pay to exchange a risky and ambiguous asset for an unambiguous asset with an identical risk and an identical expected payoff.[44] The *uncertainty premium* can be described as the total premium that a DM is willing to pay to exchange a risky and ambiguous asset for a riskless asset with an identical expected payoff. In financial markets, this premium can be labeled the *equity premium* (Brenner and Izhakian, 2018). In this view, by Theorem 2, the uncertainty premium, denoted $\mathcal{K}$, can be defined by

$$\begin{aligned} \mathrm{U}\left(\mathbb{E}\left[x\right]-\mathcal{K}\right) = & \int\limits_{x\leq k} \mathrm{U}\left(x\right)\mathrm{E}\left[\varphi\left(x\right)\right]\left(1-\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}\mathrm{Var}\left[\varphi\left(x\right)\right]\right)dx + \\ & \int\limits_{x\geq k} \mathrm{U}\left(x\right)\mathrm{E}\left[\varphi\left(x\right)\right]\left(1+\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}\mathrm{Var}\left[\varphi\left(x\right)\right]\right)dx + R_2\left(x\right), \end{aligned} \tag{10}$$

where x is the outcome of some primary act f (an asset), and $C=\mathbb{E}\left[x\right]-\mathcal{K}$ is the *certainty equivalent* outcome, satisfying $C\sim^1 f$. That is, C is the sure outcome for which the DM is willing to exchange the risky and ambiguous outcome of f. The next theorem derives the uncertainty premium and separates it into a risk premium and an ambiguity premium.

Theorem 7. *Suppose that the conditions of Theorem 6 are satisfied, and a DM with a strictly increasing and twice-differentiable utility function* U*, whose reference point k is close to the expected outcome $\mathbb{E}\left[x\right]$. When $\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}$ and $\left|x-\mathbb{E}\left[x\right]\right|$ are uncorrelated, the uncertainty premium can be written as*

$$\mathcal{K} = \underbrace{-\frac{1}{2}\frac{\mathrm{U}''\left(\mathbb{E}\left[x\right]\right)}{\mathrm{U}'\left(\mathbb{E}\left[x\right]\right)}\mathbb{V}\mathrm{ar}\left[x\right]}_{\textit{Risk Premium}}\underbrace{-\mathbb{E}\left[\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}\right]\mathbb{E}\left[\left|x-\mathbb{E}\left[x\right]\right|\right]\mho^2\left[x\right]}_{\textit{Ambiguity Premium}}+R_2\left(x\right), \tag{11}$$

where the remainder $R_2\left(x\right)=o\left(\mathbb{E}\left[\left|x-\mathbb{E}\left[x\right]\right|^2\right]\right)$ as $\left|x-\mathbb{E}\left[x\right]\right|\to 0$.[45]

Theorem 7 uses the methodology of Pratt (1964) and Arrow (1965). Hence, the accuracy of the expansion is the same as that of the Arrow-Pratt model (i.e., the remainder is of the same

[44] This premium can also be described as the price that a DM is willing to pay for information about the true probabilities of events.

[45] Namely, $\mathbb{E}\left[\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\right] = \int \mathrm{E}\left[\varphi\left(x\right)\right]\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}\left(x\right)\right]\right)}dx$ and $\mathbb{E}\left[\left|x-\mathbb{E}\left[x\right]\right|\right] = \int \mathrm{E}\left[\varphi\left(x\right)\right]\left|x-\mathbb{E}\left[x\right]\right|dx$.

order). Theorem 7 distinguishes risk and ambiguity premiums. Then, within each premium, it distinguishes the determinants of the premium: attitudes and beliefs. The risk premium is the Arrow-Pratt risk premium. However, unlike in the Arrow-Pratt settings, in which the expectation and variance are taken using known unique probabilities, here the expectation and variance are taken using the expected probabilities, as probabilities are uncertain. Independently, a higher risk, measured by $\mathbb{V}\mathrm{ar}[x]$, or a higher aversion to risk, measured by the coefficient of absolute (local) risk aversion $-\frac{\mathrm{U}''}{\mathrm{U}'}$, results in a higher risk premium.

The ambiguity premium possesses attributes that resemble those of the risk premium, but with respect to probabilities instead of outcomes. A complete separation between the extent of ambiguity, measured by $\mho^2$, and attitudes toward ambiguity, measured by the coefficient of absolute (local) ambiguity aversion $-\frac{\Upsilon''}{\Upsilon'}$, is achieved. Ambiguity aversion ($-\frac{\Upsilon''}{\Upsilon'} > 0$) implies a positive premium. Ambiguity seeking ($-\frac{\Upsilon''}{\Upsilon'} < 0$) implies a negative premium. Ambiguity indifference ($-\frac{\Upsilon''}{\Upsilon'} = 0$) implies a zero premium, which is also obtained when all probabilities are perfectly known (when $\mho^2 = 0$). A higher degree of ambiguity, or a higher aversion to ambiguity, both result in a greater ambiguity premium. The ambiguity premium is also a function of the expected absolute deviation of payoffs from their expectation, $\mathbb{E}\left[|x - \mathbb{E}[x]|\right]$. This component scales the premium to the units of outcomes. For example, when the premium is measured in terms of dollars, rather than the percentage rate of return. The next corollary demonstrates a particular case of Theorem 7.

Corollary 1. *Suppose that the conditions of Theorem 6 are satisfied and a DM with a constant absolute risk aversion and a constant absolute ambiguity aversion, whose reference point k is close to the expected outcome $\mathbb{E}[x]$. The uncertainty premium can then be written as*

$$\mathcal{K} = \underbrace{\gamma \frac{1}{2} \mathbb{V}\mathrm{ar}[x]}_{\textit{Risk Premium}} + \underbrace{\eta \mathbb{E}\left[|x - \mathbb{E}[x]|\right] \mho^2[x]}_{\textit{Ambiguity Premium}} + R_2(x),$$

where $\gamma = -\frac{\mathrm{U}''}{\mathrm{U}'}$ is the coefficient of absolute (local) risk aversion; $\eta = -\frac{\Upsilon''}{\Upsilon'}$ is the coefficient of absolute (local) ambiguity aversion; and the remainder $R_2(x) = o\left(\mathbb{E}\left[|x - \mathbb{E}[x]|^2\right]\right)$ as $|x - \mathbb{E}[x]| \to 0$.

The implications of ambiguity for the equity premium is also studied in Chen and Epstein (2002), Ui (2011), Izhakian and Benninga (2011), and Maccheroni et al. (2013). In Izhakian and Benninga (2011) and Maccheroni et al. (2013), which are based on the smooth model, the ambiguity premium is a function of the variance of the approximated certainty equivalent utilities. These, however, are a function of risk preferences, which might produce counterintuitive results. For instance, a higher risk aversion can lead to a lower uncertainty premium (Izhakian and Benninga, 2011). Generally, in previous models, the ambiguity premium is also a function of risk and attitude toward risk, whereas in Theorem 7 the ambiguity premium is independent of attitude toward risk.

Brenner and Izhakian (2018) test the pricing model in Theorem 7 using the stock trading data. Their empirical study of the risk–ambiguity–return relation aims to explain the stock market returns. It assumes that each subset of observable stock returns is generated by the choices of a representative DM, conditional upon a different prior within her subjective set of priors. Thus, the distribution of returns in each subset is used to identify the set of priors and to compute $\mho^2$, as detailed in Subsection 7.2. Brenner and Izhakian's (2018) study finds that ambiguity, as measured by $\mho^2$, has a significant impact on expected stock market returns, and that attitude toward ambiguity is contingent upon expected probabilities.

The practicality of the $\mho^2$ measure and the model in Theorem 2 is further demonstrated in other theoretical studies. For example, Izhakian et al. (2017) utilize $\mho^2$ to generalize the capital structure tradeoff theory (Modigliani and Miller, 1958) to account for ambiguity; Augustin and Izhakian (2019) utilize $\mho^2$ to study the role ambiguity plays in pricing assets in zero net supply, and show that, in contrast to risk, ambiguity may have a negative effect on the prices of options and option-like assets (CDSs, in particular); Izhakian and Zender (2018) utilize $\mho^2$ to introduce ambiguity into the principal-agent theory and show that the efficiency of contracts can be improved by inducing a specific partition of the state space into events.

7.2. Empirical implementation

To illustrate an estimation of ambiguity from the data, suppose m sample sets of outcome observations, where the number of observations in each set might be different.[46] Each sample set may be characterized by a different outcome distribution. For practical implementations, it is useful to discretize outcome distributions into n bins $B_j = \left\{s \in \mathcal{S} \mid f(s) \in \left(x_{j-1}, x_j\right]\right\}$ of equal size, such that each distribution is represented by a histogram, as demonstrated in Fig. 1.[47] The height of the bar of a particular bin is computed as the fraction (frequency) of outcomes observed in that bin, and thus represents the probability of that particular bin's outcome. As no information about the likelihood of each distribution (histogram) is available, following the principle of insufficient reason, each distribution is assigned an equal likelihood. Equipped with these m outcome histograms, the expected probability of a particular bin j across the outcome distributions, $\mathrm{E}\left[\mathrm{P}\left(B_j\right)\right]$, as well as the variance of these probabilities, $\mathrm{Var}\left[\mathrm{P}\left(B_j\right)\right]$, can then be computed. Using these values, the degree of ambiguity is then computed using the (adjusted) discrete version of $\mho^2$ in Equation (9):

$$\mho^2[f] \equiv \frac{1}{\sqrt{w(1-w)}} \sum_{j=1}^{n} \mathrm{E}\left[\mathrm{P}\left(B_j\right)\right] \mathrm{Var}\left[\mathrm{P}\left(B_j\right)\right]. \tag{12}$$

To minimize the impact of the selected bin size on the value of ambiguity, a variation of Sheppard's correction is applied to scale the weighted-average volatilities of probabilities to the bin size by $\frac{1}{\sqrt{w(1-w)}}$, where $w = x_j - x_{j-1}$.

Some bins may not be populated with outcome observations, which might make it difficult to compute their probability. To address this, assuming parametric first-order probability distributions, one can compute the parameters of the distribution of each set of outcome observations, and then use the parametric distributions to extrapolate the probability of outcomes in each bin. For example, one can compute the mean μ_i and the standard deviation σ_i of each sample set i and, assuming that outcomes follow a normal distribution (or other parametric distribution), extrapolate the missing bin probabilities by $\mathrm{P}_i\left[B_j\right] = \left[\Phi\left(x_j; \mu_i, \sigma_i\right) - \Phi\left(x_{j-1}; \mu_i, \sigma_i\right)\right]$, where

[46] For example, in computing the monthly degree of ambiguity of a given asset, each set of outcomes can be the intraday rates of return of the asset over a day. In this case, each prior in the set of priors is identified by the observed returns of the asset over a period of one day, and the number of priors in the set depends on the number of trading days in the month.

[47] Note that the set of outcome intervals induces a unique partition of the state space into events.

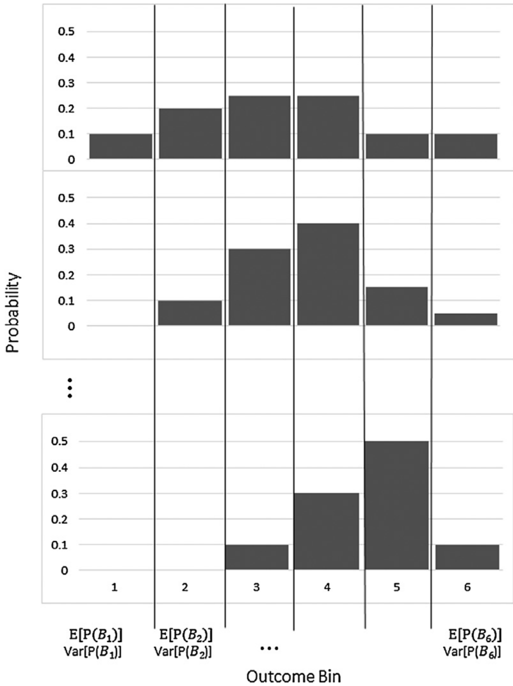

Fig. 1. Ambiguity measurement. This figure illustrates the computation of the ambiguity measure $\mho^2$. It assumes m outcome distributions in the set of priors $\mathcal{P}$. Each outcome distribution is discretized into n bins $B_j = \{s \in \mathcal{S} \mid f(s) \in (x_{j-1}, x_j]\}$ of equal size, and is represented as a histogram. The height of the bar of a particular bin is computed as the fraction of outcomes observed in that bin, and thus represents the probability of that particular bin outcome. The expected probability of a particular bin across the outcome distributions, $\mathrm{E}[\mathrm{P}(B_j)]$, as well as the variance of these probabilities, $\mathrm{Var}[\mathrm{P}(B_j)]$, can then be used to compute the degree of ambiguity as

$$\mho^2[f] = \frac{1}{\sqrt{w(1-w)}} \sum_{j=1}^{n} \mathrm{E}[\mathrm{P}(B_j)]\,\mathrm{Var}[\mathrm{P}(B_j)].$$

The weighted-average of the volatilities of probabilities is scaled to the bin size by $w = x_j - x_{j-1}$.

$\Phi(\cdot)$ denotes the cumulative normal probability distribution. Using these extrapolated probabilities, $\mathrm{E}\left[\mathrm{P}\left(E_j\right)\right]$ and $\mathrm{Var}\left[\mathrm{P}\left(E_j\right)\right]$ can be computed in order to asses the degree of ambiguity by Equation (12).

Furthermore, assuming a specific parametric form of first-order distributions, the parameters governing these first-order probability distributions, estimated from the data, can help characterize the second-order probability distribution. For example, as common in the literature, suppose that the set of priors consists of only normal distributions. Since normal distributions are characterized by a mean, μ, and a standard deviation, σ, once these two parameters are extracted, their joint distribution can characterize the second-order probability distribution. This distribution may be interpreted as the second-order belief of a representative DM.

This method of estimating ambiguity from the data is applied in several empirical studies that estimate the ambiguity associated with the stock market and with individual firms. These studies find that the $\mho^2$ measure helps explain financial decisions and pricing; such as, capital structure decisions (Izhakian et al., 2017); exercising executive options (Izhakian and Yermack, 2017);

pricing credit default swaps (Augustin and Izhakian, 2019); firms' payout policies (Herron and Izhakian, 2018); market equity premium (Brenner and Izhakian, 2018); and trading activities around macroeconomic announcements (Doan et al., 2018).

8. Conclusion

Almost any financial and economic decision entails ambiguity. Consciously or subconsciously, one of the first steps in a decision-making process is to rank alternative choices by their degree of ambiguity. The importance of ambiguity for understanding financial and economic decision-making has been recognized in the literature for the past century. Relevant studies have acknowledged that any attempt to portray a realistic picture of observable phenomena must also consider the uncertainty about probabilities. A simple, well-defined, risk-independent measure of the extent of ambiguity is the key to account for this dimension of uncertainty. The search for a measure that can quantify the degree of ambiguity associated with different alternatives was initiated with the seminal study of Knight (1921). The measure of ambiguity introduced in this paper aims to advance that search.

By describing ambiguity as probability perturbations (uncertain probabilities) and aversion to ambiguity as aversion to the mean-preserving spreads in these probabilities, this paper introduces a natural *ambiguity measure*, denoted $\mho^2$, which is simply the expected volatility of probabilities across the relevant events. This measure has two main qualities: first, it is simple, intuitive, and empirically applicable for measuring the degree of ambiguity; second, it is *outcome independent* (up to a state space partition) and, therefore, *risk independent*. Moreover, the $\mho^2$ measure is independent of individuals' attitudes toward ambiguity and risk. These qualities are critical when introducing ambiguity into theoretical, behavioral, and empirical studies, and are necessary for studying the distinct role that the extent of ambiguity plays in decision-making processes.

Almost a century ago, Knight (1921) recognized the need for an applicable, risk-independent measure of the extent of ambiguity. While advancing the literature toward addressing this need, the measure of ambiguity introduced in this paper may be further developed to support broader settings. The concept of an outcome-independent ambiguity measurement, introduced in this paper, may also stimulate further thinking that will advance the literature toward a better understanding of ambiguity.

Appendix A

A.1. Expected utility with uncertain probabilities

For completeness, this section formally presents and discusses the EUUP model assumptions. The EUUP model assumes the following:

Assumption 1. For every $c \in [0, 1]$, there exists an event $E \in \mathcal{H}$ with the probability $\mathrm{P}(E) = c$.[48]

[48] Note that this assumption is less restrictive than an alternative assumption of a state space $\mathcal{S} \times (0, 1]$, endowed with a product σ-algebra of $\mathcal{E}$ and a Borel σ-algebra of $(0, 1]$.

Assumption 2. The preference relation $\succsim^1$ over the set of primary acts $\mathcal{F}_0$ satisfies the CPT axioms of Wakker (Theorem 12.3.5, 2010).[49]

Assumption 3. For any $E, F \in \mathcal{E}$, $\delta_E \succsim^1 \delta_F \iff -\delta_{E^C} \succsim^1 -\delta_{F^C}$.[50]

Assumption 4. The preference relation $\succsim^2$ over the set of secondary acts $\widehat{\mathcal{F}}$ satisfies the extended Savage axioms of Kopylov (2010), and $x > y \iff x \succ^2 y$, for all $x, y \in X$.[51]

Assumption 5. Jointly, $\succsim^1$ and $\succsim^2$ satisfy the certainty equivalent probabilistic consistency axiom (Axiom 1) of Izhakian (2017).

Under these assumptions, preferences for ambiguity, $\succsim^2$, in the EUUP model, are outcome independent. As a frame of reference, consider the smooth model (Klibanoff et al., 2005). Superficially, the EUUP model proposed in Equation (2) resembles the smooth model. However, although both models rely on the idea of second-order preferences (preferences for ambiguity) over secondary acts, they are conceptually different. In the smooth model, second-order preferences apply to expected (or certainty equivalent) utilities, which are subject to risk and preferences for risk. Therefore, in the smooth model, preferences for ambiguity are outcome dependent, risk dependent, and attitude (toward risk) dependent.

To avoid these dependencies and refine the separations between risk and ambiguity, in the EUUP model, the consistency of first-order preference, $\succsim^1$, with second-order preference, $\succsim^2$, is less restrictive. In the smooth model (Klibanoff et al., 2005, Assumption 3), $\succsim^1$ satisfies $f \succsim^1 g \iff f^2 \succsim^2 g^2$ for all acts, where $f^2(\mathrm{P})$ is the certainty equivalent of f computed using $\mathrm{P} \in \mathcal{P}$.[52] Thus, the source of the preference relation $f^2 \succsim^2 g^2$ cannot be determined explicitly, since f^2 is outcome-dependent, and subject to risk and attitude toward risk; therefore, $\succsim^2$ cannot be attributed exclusively to ambiguity. On the other hand, in the EUUP model (Izhakian, 2017, Axiom 1 and Proposition 1), $\succsim^1$ is required to satisfy $\delta_E \succsim^1 \delta_F \iff \hat{\delta}_E \succsim^2 \hat{\delta}_F$ only over primary indicator acts. The less restrictive consistency axiom allows preferences for ambiguity, $\succsim^2$, to be outcome independent, and to apply directly and exclusively to probabilities (rather than to certainty equivalent utilities). This allows to define preferences for ambiguity as preferences for mean-preserving spreads in probabilities. With this characteristic in hand, preferences for ambi-

[49] Alternative sets of CPT axioms can be adopted (e.g., Tversky and Kahneman, 1992; Wakker and Tversky, 1993; Hong and Wakker, 1996; Kothiyal et al., 2011). When considering a DM whose preferences are reference independent, the CEU axioms of Schmeidler (1989), for example, can also be adopted.

[50] Allowing the impact of an event on the desirability of the act ("decision weight") to depend on the outcome associated with the event, this assumption can be dropped.

[51] The EUUP model employs the Kopylov (2010) axioms instead of the Savage (1954) axioms, since secondary acts do not generally have a finite range, and evaluating their expected utility would generally require a countably-additive measure. Alternative sets of axioms can be adopted, as long as they provide a unique up to a positive linear transformation utility function and a unique countably-additive probability measure (e.g., Wakker, 1993; Kopylov, 2007).

[52] In the smooth model, preferences for ambiguity are captured by a function over expected utilities, where each expected utility corresponds to a $\mathrm{P} \in \mathcal{P}$, i.e., an "expected utility over expected utilities." Formally, in the smooth model, the value of a primary act is assessed by $\mathrm{V}(f) = \int_{\mathcal{P}} \phi \left(\int_{\mathcal{S}} \mathrm{U}(f(s))\, d\mathrm{P} \right) d\xi$, where ϕ captures attitudes toward ambiguity.

guity in the EUUP model facilitate an outcome-independent (up to a state space partition), and a risk-independent measure of ambiguity.

A.2. Lemmata

Lemma 1. *Let event $E^C \in \mathcal{E}$ be the complementary of event $E \in \mathcal{E}$. The covariance of their probabilities,* $\mathrm{Cov}\left[\mathrm{P}(E), \mathrm{P}(E^C)\right] \equiv \int_{\mathcal{P}} \left(\mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]\right)\left(\mathrm{P}(E^C) - \mathrm{E}\left[\mathrm{P}(E^C)\right]\right) d\xi$, *then satisfies*

$$-\mathrm{Cov}\left[\mathrm{P}(E), \mathrm{P}(E^C)\right] = \mathrm{Var}[\mathrm{P}(E)] = \mathrm{Var}\left[\mathrm{P}(E^C)\right].$$

Lemma 2. *Suppose that* $\{\mathrm{P}(E)\}_{\mathrm{P}\in\mathcal{P}}$ *and* $\{\mathrm{P}(F)\}_{\mathrm{P}\in\mathcal{P}}$ *are uniformly distributed, or truncated elliptically distributed with an identical characteristic generator, and have an identical expectation. Let* $\mathrm{Std}[\mathrm{P}(E)]$ *and* $\mathrm{Std}[\mathrm{P}(F)]$ *be, respectively, the standard deviations of* $\{\mathrm{P}(E)\}_{\mathrm{P}\in\mathcal{P}}$ *and* $\{\mathrm{P}(F)\}_{\mathrm{P}\in\mathcal{P}}$. *Then,*

$$\mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)] =_d \lambda\left(\mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]\right),$$

where $\lambda = \frac{\mathrm{Std}[\mathrm{P}(F)]}{\mathrm{Std}[\mathrm{P}(E)]}$.

Lemma 3. *Let $h : \mathbb{R} \to \mathbb{R}$ be some function. The volatility of the probabilities of x is then uncorrelated with $h(x)$, for any $a \le x \le b$. That is,* $\mathbb{C}\mathrm{ov}\left[h(x), \mathrm{Var}[\varphi(x)] \mid a \le x \le b\right] = 0$,[53] *implying that*

$$\mathbb{E}\left[h(x)\mathrm{Var}[\varphi(x)] \mid a \le x \le b\right] = \mathbb{E}\left[h(x) \mid a \le x \le b\right] \mathbb{E}\left[\mathrm{Var}[\varphi(x)] \mid a \le x \le b\right].$$

A.3. Proofs

Proof of Lemma 1. Immediately obtained, since every $\mathrm{P} \in \mathcal{P}$ is additive. □

Proof of Lemma 2. Let $y = \mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]$ and $z = \mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)]$. Suppose that y and z are uniformly distributed. Since $\mathrm{E}[y] = 0$ and $\mathrm{E}[z] = 0$, every $y \in \left[-a_y, a_y\right]$ and every $z \in \left[-a_z, a_z\right]$. The characteristic function of z and λy are, respectively,

$$\phi_z(t) = \frac{e^{ita_z} - e^{-ita_z}}{2ita_z} \qquad \text{and} \qquad \phi_{\lambda y}(t) = \frac{e^{it\lambda a_y} - e^{-it\lambda a_y}}{2it\lambda a_y}.$$

Since y and z are uniformly distributed with $\mathrm{E}[z] = \mathrm{E}[y] = 0$, $\lambda = \frac{\mathrm{Std}[z]}{\mathrm{Std}[y]} = \sqrt{\frac{(2a_z)^2/12}{\left(2a_y\right)^2/12}}$, which implies $a_z = \lambda a_y$. Thus, $\phi_z(t) = \phi_{\lambda y}(t)$ for any t, and therefore $z =_d \lambda y$.

Suppose that z and λy are truncated elliptically distributed. Their characteristic functions are

$$\phi_z(t) = e^{it\mathrm{E}[z]}\Psi\left(\frac{1}{2}t^2\mathrm{Var}[z]\right) \qquad \text{and} \qquad \phi_{\lambda y}(t) = e^{it\lambda\mathrm{E}[y]}\Psi\left(\frac{1}{2}t^2\lambda^2\mathrm{Var}[y]\right),$$

respectively. Since ϕ_z and $\phi_{\lambda y}$ have an identical characteristic generator Ψ, $\mathrm{E}[z] = \lambda\mathrm{E}[y] = 0$, and $\mathrm{Std}[z] = \lambda\mathrm{Std}[y]$, then $\phi_z = \phi_{\lambda y}$, which implies that $z =_d \lambda y$. □

---

[53] $\mathbb{C}\mathrm{ov}\left[\cdot\right]$ stands for the covariance of outcomes, taken using the expected probabilities.

Proof of Lemma 3. Let $y = \varphi(x)$, then $\mathrm{Var}[\varphi(x)]$ can be written as $\mathrm{Var}[y|x] = \mathrm{E}[y^2|x] - \mathrm{E}^2[y|x]$. In turn, $C = \mathbb{C}\mathrm{ov}\left[\mathrm{Var}[\varphi(x)], h(x) \mid a \le x \le b\right]$ can be written explicitly as

$$\begin{aligned} C &= \mathbb{E}\left[\left(\mathrm{E}[y^2|x] - \mathrm{E}^2[y|x] - \mathbb{E}\left[\mathrm{E}[y^2|x] - \mathrm{E}^2[y|x] \,\middle|\, a \le x \le b\right]\right)\right. \\ &\qquad \left.\times \left(h(x) - \mathbb{E}[h(x) \mid a \le x \le b]\right) \,\middle|\, a \le x \le b\right] \\ &= \mathbb{E}\left[h(x)\left(\mathrm{E}[y^2|x] - \mathrm{E}^2[y|x]\right) \mid a \le x \le b\right] \\ &\qquad - \mathbb{E}\left[h(x)\mathbb{E}\left[\mathrm{E}(y^2|x) - \mathrm{E}^2[y|x] \mid a \le x \le b\right] \mid a \le x \le b\right]. \end{aligned}$$

Applying the tower property to the first term, and the law of iterated expectations to the second term (e.g., Goldberger, 1991, p. 47, T8), provides

$$\begin{aligned} C &= \mathbb{E}\left[h(x)\left(\mathrm{E}[y^2] - \mathrm{E}^2[y]\right) \mid a \le x \le b\right] - \mathbb{E}\left[h(x)\mathbb{E}\left[\mathrm{E}[y^2]\right.\right. \\ &\qquad \left.\left. - \mathrm{E}^2[y] \mid a \le x \le b\right] \mid a \le x \le b\right]. \end{aligned}$$

By Karlin and Taylor (2012, p. 8), $\mathbb{E}\left[h(x)\mathbb{E}[g(y)|x]\right] = \mathbb{E}\left[h(x)g(y)\right]$. Therefore,

$$\begin{aligned} &\mathbb{E}\left[h(x)\mathbb{E}\left[\mathrm{E}[y^2] - \mathrm{E}^2[y] \mid a \le x \le b\right] \mid a \le x \le b\right] \\ &\quad = \mathbb{E}\left[h(x)\left(\mathrm{E}[y^2] - \mathrm{E}^2[y]\right) \mid a \le x \le b\right], \end{aligned}$$

which completes the proof. □

Proof of Theorem 1. Omitting the subscript f, the marginal perceived probability in Equation (4) is

$$q = \frac{d}{dx}\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1 - \mathrm{P}(x))]\right) = -\frac{\mathrm{E}\left[\Upsilon'(1 - \mathrm{P}(x))\varphi(x)\right]}{\Upsilon'\left(\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1 - \mathrm{P}(x))]\right)\right)}. \tag{A.1}$$

Notice that, since $\varphi(t)$ is additive, by changing the integration order, $\int_{-\infty}^{x} \mathrm{E}[\varphi(t)]\,dt = \mathrm{E}[\mathrm{P}(x)]$.

Let $G = \frac{\Upsilon'(1-\mathrm{P}(x))}{\Upsilon'\left(\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1-\mathrm{P}(x))]\right)\right)}$. By Judd (2003), the first-order Taylor expansion of G around $\mathrm{E}[\varphi(x)]$ is[54,55]

[54] Judd (2003) shows that the Taylor expansion of $f(x)$ can be improved by the change of variable $x = h(y)$, i.e., writing x as a non-linear transformation of y, to obtain h-linearization and expanding $f(h(y))$ with respect to y. Here, the linearization is applied by $\varphi(\cdot)$.

[55] Note that $\frac{d}{dx}\Upsilon'\left(\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1 - \mathrm{P}(x))]\right)\right) = -\frac{\Upsilon''\left(\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1-\mathrm{P}(x))]\right)\right)\mathrm{E}\left[\Upsilon'(1-\mathrm{P}(x))\varphi(x)\right]}{\Upsilon'\left(\Upsilon^{-1}\left(\mathrm{E}[\Upsilon(1-\mathrm{P}(x))]\right)\right)}$.

$$G = 1 - \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)](\varphi(x)-\mathrm{E}[\varphi(x)]) + \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\mathrm{E}\big[\varphi(x)-\mathrm{E}[\varphi(x)]\big] + R_{G,2}(\varphi(x)). \tag{A.2}$$

Now, the marginal perceived probability q in Equation (A.1) can be expanded by

$$\begin{aligned} q &= -\mathrm{E}[G\varphi(x)] \\ &= -\mathrm{E}[\varphi(x)] + \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\mathrm{E}\big[\big(\varphi(x) \\ &\quad - \mathrm{E}[\varphi(x)]\big)\varphi(x)\big] - \mathrm{E}\left[R_{G,2}(\varphi(x))\varphi(x)\right] \\ &= -\mathrm{E}[\varphi(x)] + \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\mathrm{E}\big[\big(\varphi(x)-\mathrm{E}[\varphi(x)]\big)^2\big] - R_2(\varphi(x)), \end{aligned}$$

where $R_2(\varphi(x))$ stands for $\mathrm{E}\left[R_{G,2}(\varphi(x))\varphi(x)\right]$. Therefore,

$$q = -\mathrm{E}[\varphi(x)] + \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\,\mathrm{Var}[\varphi(x)] - R_2(\varphi(x)). \tag{A.3}$$

Substituting q for the marginal perceived probabilities in Equation (4), while accounting for the sign switch of $\mathrm{E}[\varphi(x)]$ when moving from a negative to a positive utility across k (Wakker and Tversky, 1993),[56] provides Equation (5).

The remainder in Equation (A.2) is

$$\begin{aligned} R_{G,2}(\varphi(x)) = &\frac{\Upsilon'(1-\mathrm{P}(x))}{\Upsilon'\big(\Upsilon^{-1}\big(\mathrm{E}[\Upsilon(1-\mathrm{P}(x))]\big)\big)} \\ &- 1 + \frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)](\varphi(x)-\mathrm{E}[\varphi(x)]). \end{aligned}$$

Let $\Delta = \varphi(x) - \mathrm{E}[\varphi(x)]$, then

$$\lim_{\Delta\to 0}\frac{R_{G,2}(\Delta+\mathrm{E}[\varphi(x)])}{\Delta^2} = 0.$$

To prove this, let $A = G - R_{G,2}(\varphi(x))$ be the approximated term in Equation (A.2). Then, G and G' agree with A and A' at $\Delta = 0$. By Lagrange's remainder theorem, $R_{G,2}(\Delta+\mathrm{E}[\varphi(x)]) = \frac{G''(c)}{2!}\Delta^2$ for some $c \in (0, \Delta+\mathrm{E}[\varphi(x)])$. Thus, $R_{G,2}$ and $R'_{G,2}$ vanish as $\Delta \to 0$. Moreover, since Δ^2 and its first two derivatives also vanish as $\Delta \to 0$, the l'Hôpital rule can be applied twice to obtain the following

$$\lim_{\Delta\to 0}\frac{R_{G,2}(\Delta+\mathrm{E}[\varphi(x)])}{\Delta^2} = \lim_{\Delta\to 0}\frac{R''_{G,2}(\Delta+\mathrm{E}[\varphi(x)])}{2\cdot 1\cdot \Delta^0} = \lim_{\Delta\to 0}\frac{0}{2!} = 0.$$

By Equation (A.2), $R_2 = \mathrm{E}\left[\frac{G''(c)}{2!}\Delta^2\varphi(x)\right] = \frac{G''(c)}{2!}\mathrm{E}\left[\Delta^3\right] + \frac{G''(c)}{2!}\mathrm{E}\left[\Delta^2\right]\mathrm{E}[\varphi(x)]$. By similar considerations, applying the l'Hôpital rule three times gives

56 By Wakker and Tversky (1993), the sign switch is determined by a linear shift, which ensures that capacities (perceived probabilities) are nonnegative. This can also be viewed through the Choquet integration over negative functions, which takes the form $\int f\,dQ = \int (f+c)\,dQ - c$, where $c > 0$ such that $f + c > 0$.

$$\lim_{\Delta\to 0}\frac{R_2(\Delta+\mathrm{E}[\varphi(x)])}{\mathrm{E}\left[\Delta^3\right]} = \lim_{\Delta\to 0}\frac{R_2'''(\Delta+\mathrm{E}[\varphi(x)])}{3\cdot 2\cdot 1\cdot \mathrm{E}\left[\Delta^0\right]} = \lim_{\Delta\to 0}\frac{0}{3!} = 0.$$

Therefore, $R_2(\varphi(x)) = o\left(\mathrm{E}\left[|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]\right)$ as $|\varphi(x)-\mathrm{E}[\varphi(x)]| \to 0$. □

Proof of Theorem 2. Omitting the subscript f and substituting the marginal perceived probabilities, expanded in Equation (5), into Equation (4) gives

$$\mathrm{W}(f) = \int\limits_{x\le k} \mathrm{U}(x)\left(\mathrm{E}[\varphi(x)] - \tfrac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\operatorname{Var}[\varphi(x)] + R_2(\varphi(x))\right)dx +$$
$$\int\limits_{x\ge k} \mathrm{U}(x)\left(\mathrm{E}[\varphi(x)] + \tfrac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\operatorname{Var}[\varphi(x)] + R_2(\varphi(x))\right)dx.$$

To show that the remainder $R_2(f) = \int \mathrm{U}(x)\,R_2(\varphi(x))\,dx$ vanishes faster than $\int \mathrm{E}\big[|\varphi(x) - \mathrm{E}[\varphi(x)]|^3\big]x\,dx$ as $\int |\varphi(x)-\mathrm{E}[\varphi(x)]|\,dx \to 0$, write

$$\lim_{t\to 0}\frac{\int \mathrm{U}(x)\,R_2(t\varphi(x))\,dx}{\int \mathrm{E}\left[t^3|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]x\,dx} = \lim_{t\to 0}\frac{\int \mathrm{U}(x)\,R_2'''(t\varphi(x))\,\varphi^3(x)\,dx}{3\cdot 2\cdot 1\cdot t^0\int \mathrm{E}\left[|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]x\,dx}$$
$$= \lim_{t\to 0}\frac{0}{3!\int \mathrm{E}\left[|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]x\,dx} = 0,$$

where the first equality is obtained by applying the l'Hôpital rule three times, and the second by the same considerations applied to $R_2(\varphi(x))$ in the proof of Theorem 1. □

Proof of Theorem 3. Let $y = \mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]$ and $z = \mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)]$, and F be more ambiguous than E. By Definition 4, $z =_d y + \epsilon$. Thus,

$$\mathrm{E}[\Upsilon(z)] = \mathrm{E}\big[\mathrm{E}[\Upsilon(y+\epsilon)\mid y]\big].$$

Ignoring the first expectation on the right-hand side for the moment, a concave Υ (aversion to ambiguity) implies

$$\mathrm{E}[\Upsilon(z)] = \mathrm{E}[\Upsilon(y+\epsilon)] \le \Upsilon(\mathrm{E}[y+\epsilon]) = \Upsilon(y).$$

Taking expectation implies $\mathrm{E}[\Upsilon(z)] \le \mathrm{E}[\Upsilon(y)]$. Hence, by Izhakian (2017, Proposition 2), $\hat{\delta}_F \precsim^2 \hat{\delta}_E$.

For the opposite direction, let $\hat{\delta}_F \precsim^2 \hat{\delta}_E$. Then, by Izhakian (2017, Proposition 2), $\mathrm{E}[\Upsilon(z)] \le \mathrm{E}[\Upsilon(y)]$. To complete the proof, the existence of ϵ that satisfies $z =_d y + \epsilon$ (Definition 4) has to be proven. This proof considers two probability measures $\mathrm{P} \in \mathcal{P}$; it can then be extended to any number of measures. Let y and z take two possible values, (y_1, y_2) and (z_1, z_2), with probabilities $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$, respectively. Without loss of generality, assume that $z_1 \ge y_1 \ge y_2 \ge z_2$. Then, ϵ can be constructed as $\epsilon_1 = (z_1 - y_1, z_2 - y_1)$ with probabilities $\left(\frac{y_1-z_2}{z_1-z_2}, \frac{z_1-y_1}{z_1-z_2}\right)$ and $\epsilon_2 = (z_1 - y_2, z_2 - y_2)$ with probabilities $\left(\frac{y_2-z_2}{z_1-z_2}, \frac{z_1-y_2}{z_1-z_2}\right)$. The probabilities of ϵ_1 and ϵ_2 are all positive and less than 1, and $\mathrm{E}[\epsilon_1 \mid y_1] = 0$ and $\mathrm{E}[\epsilon_2 \mid y_2] = 0$. Therefore, ϵ is mean independent of y and $\mathrm{E}[z] = \mathrm{E}[y+\epsilon] = 0$. The probability that $y+\epsilon = z_1$ is $\alpha\frac{y_1-z_2}{z_1-z_2} + (1-\alpha)\frac{y_2-z_2}{z_1-z_2}$. Since $\mathrm{E}[y] = \mathrm{E}[z]$, then $\alpha = \frac{z_2-y_2+\beta(z_1-z_2)}{y_1-y_2}$. Together, this implies that the probability that $y+\epsilon = z_1$ is equal to β, and the probability that $y+\epsilon = z_2$ is equal to $1-\beta$. That is, $z =_d y+\epsilon$.

Izhakian (2017, Proposition 1) completes the proof. □

Proof of Theorem 4. Suppose ambiguity aversion (the proof for ambiguity seeking is similar). $(i+ii)$ Let $y = \mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]$ and $z = \mathrm{P}(F) - \mathrm{E}[\mathrm{P}(F)]$, and F be more ambiguous than E. By Definition 4, $z =_d y+\epsilon$, where ϵ is mean independent of y. By Goldberger (1991, p. 63, M2), mean independence implies uncorrelatedness; therefore,

$$\mathrm{Var}[\mathrm{P}(F)] = \mathrm{Var}[\mathrm{P}(E)] + \mathrm{Var}[\epsilon].$$

For the opposite direction, let $\mathrm{Var}[\mathrm{P}(F)] \geq \mathrm{Var}[\mathrm{P}(E)]$ and $\lambda = \frac{\mathrm{Std}[\mathrm{P}(F)]}{\mathrm{Std}[\mathrm{P}(E)]}$. Since y and z are either uniformly distributed or truncated elliptically distributed with an identical characteristic generator, and $\mathrm{E}[z] = \mathrm{E}[y]$, by Lemma 2, $z =_d \lambda y$. Write $x+y = \alpha(x+\lambda y) + (1-\alpha)x$, where $\alpha = \frac{1}{\lambda}$ and x is a random variable that satisfies $\mathrm{E}[x \mid y] = \mathrm{E}[x]$. Since Υ is concave (aversion to ambiguity), by the Jensen inequality,

$$\Upsilon(x+y) \geq \alpha\Upsilon(x+\lambda y) + (1-\alpha)\Upsilon(x).$$

Taking expectations of both sides provides

$$\mathrm{E}\big[\mathrm{E}[\Upsilon(x+y) \mid x]\big] \geq \alpha\mathrm{E}\big[\mathrm{E}[\Upsilon(x+\lambda y) \mid x]\big] + (1-\alpha)\mathrm{E}[\Upsilon(x)]. \tag{A.4}$$

Since $\mathrm{E}[\lambda y] = 0$, a concave Υ implies

$$\mathrm{E}\big[\Upsilon(\mathrm{E}[x+\lambda y \mid x])\big] = \mathrm{E}[\Upsilon(x)] \geq \mathrm{E}\big[\mathrm{E}[\Upsilon(x+\lambda y) \mid x]\big],$$

which jointly with Equation (A.4) imply

$$\mathrm{E}\big[\mathrm{E}[\Upsilon(x+y) \mid x]\big] \geq \mathrm{E}\big[\mathrm{E}[\Upsilon(x+\lambda y) \mid x]\big].$$

In particular, this holds true for $x = 0$; hence,

$$\mathrm{E}[\Upsilon(y)] \geq \mathrm{E}[\Upsilon(\lambda y)] = \mathrm{E}[\Upsilon(z)].$$

(iii) Let $\Upsilon(\mathrm{P}(E)) = \frac{1-e^{-\eta\mathrm{P}(E)}}{\eta}$. The second-order Taylor expansion around $\mathrm{E}[\mathrm{P}(E)]$ is

$$\begin{aligned}\Upsilon(\mathrm{P}(E)) = \frac{1}{\eta} - \frac{e^{-\eta\mathrm{E}[\mathrm{P}(E)]}}{\eta}\left(1 - \eta\big(\mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]\big) + \frac{1}{2}\eta^2\big(\mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]\big)^2\right)\\ + o\left(|\mathrm{P}(E) - \mathrm{E}[\mathrm{P}(E)]|^2\right).\end{aligned}$$

Taking expectation provides

$$\mathrm{E}[\Upsilon(\mathrm{P}(E))] = \frac{1}{\eta} - \frac{e^{-\eta\mathrm{E}[\mathrm{P}(E)]}}{\eta}\left(1 + \frac{1}{2}\eta^2\mathrm{Var}[\mathrm{P}(E)]\right) + o\big(\mathrm{Var}[\mathrm{P}(E)]\big).$$

Since $\mathrm{P}(E)$ is close to $\mathrm{E}[\mathrm{P}(E)]$, the third and higher absolute central moments of $\mathrm{P}(E)$ are negligible. Thus, since $\mathrm{E}[\mathrm{P}(E)] = \mathrm{E}[\mathrm{P}(F)]$, concerning ambiguity aversion ($\eta > 0$),

$$\mathrm{E}[\Upsilon(\mathrm{P}(E))] \geq \mathrm{E}[\Upsilon(\mathrm{P}(F))] \iff \mathrm{Var}[\mathrm{P}(E)] \leq \mathrm{Var}[\mathrm{P}(F)].$$

(iv) Let $\Upsilon(\mathrm{P}(E)) = -\big(\mathrm{P}(E) - \alpha\big)^2$, where $\mathrm{P}(E) \leq \alpha$ for some $\alpha \in \mathbb{R}$. Taking expectation provides

$$\mathrm{E}[\Upsilon(\mathrm{P}(E))] = -\mathrm{Var}[\mathrm{P}(E)] - \left(\mathrm{E}[\mathrm{P}(E)] - \alpha\right)^2.$$

Since $\mathrm{E}[\mathrm{P}(E)] = \mathrm{E}[\mathrm{P}(F)]$,

$$\mathrm{E}[\Upsilon(\mathrm{P}(E))] \geq \mathrm{E}[\Upsilon(\mathrm{P}(F))] \Longleftrightarrow \mathrm{Var}[\mathrm{P}(E)] \leq \mathrm{Var}[\mathrm{P}(F)].$$

Together, Izhakian (2017, Proposition 2) and Theorem 3, complete the proof. □

Proof of Theorem 5. Since $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_g(x)\right]$ for any $x \in X$ and reference dependence is dropped, by Theorem 2,

$$\begin{aligned}\mathrm{W}(f) - \mathrm{W}(g) = \int \mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} \mathrm{E}\left[\varphi_f(x)\right] \big(&\mathrm{Var}\left[\varphi_f(x)\right] \\ &- \mathrm{Var}\left[\varphi_g(x)\right]\big) dx + R_2(f-g).\end{aligned}$$

Since, by Lemma 3, $\mathrm{Var}\left[\varphi_f(x)\right]$ and $\mathrm{Var}\left[\varphi_g(x)\right]$ are both uncorrelated with $\mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}$,

$$\begin{aligned}\mathrm{W}(f) - \mathrm{W}(g) = \int \mathrm{E}\left[\varphi_f(x)\right] \mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} dx \int \mathrm{E}\left[\varphi_f(x)\right] \big(&\mathrm{Var}\left[\varphi_f(x)\right] \\ &- \mathrm{Var}\left[\varphi_g(x)\right]\big) dx + R_2(f-g).\end{aligned}$$

By aversion to ambiguity, $\int \mathrm{E}\left[\varphi_f(x)\right] \mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} dx \leq 0$. Thus, since by Theorem 2, $R_2(f-g) = o\left(\int \mathrm{E}\left[|\varphi(x) - \mathrm{E}[\varphi(x)]|^3\right] x dx\right)$ is negligible,

$$\mathrm{W}(f) \geq \mathrm{W}(g) \iff \int \mathrm{E}\left[\varphi_f(x)\right] \mathrm{Var}\left[\varphi_f(x)\right] dx \leq \int \mathrm{E}\left[\varphi_f(x)\right] \mathrm{Var}\left[\varphi_g(x)\right] dx.$$

By Theorem 2,

$$f \succsim^1 g \iff \mho^2[f] \leq \mho^2[g]. \quad \square$$

Proof of Theorem 6. Since $\mathrm{E}\left[\varphi_f(x)\right] = \mathrm{E}\left[\varphi_g(x)\right]$ for any $x \in X$, by Theorem 2,

$$\begin{aligned}\mathrm{W}(f) - \mathrm{W}(g) = &- \int_{x \leq k} \mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} \mathrm{E}\left[\varphi_f(x)\right] \big(\mathrm{Var}\left[\varphi_f(x)\right] \\ &- \mathrm{Var}\left[\varphi_g(x)\right]\big) dx \\ &+ \int_{x \geq k} \mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)} \mathrm{E}\left[\varphi_f(x)\right] \big(\mathrm{Var}\left[\varphi_f(x)\right] \\ &- \mathrm{Var}\left[\varphi_g(x)\right]\big) dx + R_2(f-g).\end{aligned}$$

By Lemma 3, $\mathrm{Var}\left[\varphi_f(x)\right]$ and $\mathrm{Var}\left[\varphi_g(x)\right]$ are both uncorrelated with $\mathrm{U}(x) \frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}$ over any subset of X. Thus,

$$
\begin{aligned}
\mathrm{W}(f)-\mathrm{W}(g)= & -\int_{x\le k}\mathrm{E}\left[\varphi_f(x)\right]\mathrm{U}(x)\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}dx\\
& \times\int_{x\le k}\mathrm{E}\left[\varphi_f(x)\right]\left(\mathrm{Var}\left[\varphi_f(x)\right]-\mathrm{Var}\left[\varphi_g(x)\right]\right)dx\\
& +\int_{x\ge k}\mathrm{E}\left[\varphi_f(x)\right]\mathrm{U}(x)\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}dx\\
& \times\int_{x\ge k}\mathrm{E}\left[\varphi_f(x)\right]\left(\mathrm{Var}\left[\varphi_f(x)\right]-\mathrm{Var}\left[\varphi_g(x)\right]\right)dx+R_2(f-g).
\end{aligned}
$$

Aversion to ambiguity and reference-dependent U imply $-\int_{x\le k}\mathrm{E}\left[\varphi_f(x)\right]\mathrm{U}(x)\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}dx \le 0$ and $\int_{x\ge k}\mathrm{E}\left[\varphi_f(x)\right]\mathrm{U}(x)\frac{\Upsilon''\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}{\Upsilon'\left(1-\mathrm{E}\left[\mathrm{P}_f(x)\right]\right)}dx\le 0$. Thus, because outcomes are symmetrically distributed around k, and by Theorem 2 $R_2(f-g)=o\left(\int\mathrm{E}\left[|\varphi(x)-\mathrm{E}[\varphi(x)]|^3\right]x\,dx\right)$ is negligible,

$$
\begin{aligned}
\mathrm{W}(f)-\mathrm{W}(g)\ge 0 \iff & \int_{x\le k}\mathrm{E}\left[\varphi_f(x)\right]\left(\mathrm{Var}\left[\varphi_f(x)\right]-\mathrm{Var}\left[\varphi_g(x)\right]\right)dx+\\
& \int_{x\ge k}\mathrm{E}\left[\varphi_f(x)\right]\left(\mathrm{Var}\left[\varphi_f(x)\right]-\mathrm{Var}\left[\varphi_g(x)\right]\right)dx \le 0,
\end{aligned}
$$

which (again by symmetry) implies

$$
\mathrm{W}(f)\ge\mathrm{W}(g) \iff \int\mathrm{E}\left[\varphi_f(x)\right]\mathrm{Var}\left[\varphi_f(x)\right]dx\le\int\mathrm{E}\left[\varphi_f(x)\right]\mathrm{Var}\left[\varphi_g(x)\right]dx.
$$

Finally, by Theorem 2,

$$
f\succsim^1 g \iff \mho^2[f]\le\mho^2[g]. \quad \square
$$

Proof of Theorem 7. The first-order Taylor expansion of the left-hand side (LHS) of Equation (10) with respect to $\mathcal{K}$, around 0, is

$$
LHS=\mathrm{U}(\mathbb{E}[x]-\mathcal{K})=\mathrm{U}(\mathbb{E}[x])-\mathcal{K}\mathrm{U}'(\mathbb{E}[x])+o(|x|).
$$

Writing the right-hand side (RHS) of Equation (10) as

$$
\begin{aligned}
RHS= & \int\mathrm{E}[\varphi(x)]\mathrm{U}(x)\,dx-\int_{x\le k}\mathrm{U}(x)\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\mathrm{Var}[\varphi(x)]\,dx \qquad \text{(A.5)}\\
& +\int_{x\ge k}\mathrm{U}(x)\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\mathrm{E}[\varphi(x)]\mathrm{Var}[\varphi(x)]\,dx+R_2(x),
\end{aligned}
$$

the second-order Taylor expansion of the first component with respect to x, around $\mathbb{E}[x]$, is

$$
\begin{aligned}
I = &\int \mathrm{E}\left[\varphi(x)\right]\Big(\mathrm{U}\left(\mathbb{E}\left[x\right]\right) + \mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\left(x - \mathbb{E}\left[x\right]\right) + \frac{1}{2}\mathrm{U}''\left(\mathbb{E}\left[x\right]\right)\left(x - \mathbb{E}\left[x\right]\right)^2 \\
&+ o\left(\left|x - \mathbb{E}\left[x\right]\right|^2\right)\Big)dx \\
= &\,\mathrm{U}\left(\mathbb{E}\left[x\right]\right) + \frac{1}{2}\mathrm{U}''\left(\mathbb{E}\left[x\right]\right)\mathbb{V}\mathrm{ar}\left[x\right] + o\left(\mathbb{E}\left[\left|x - \mathbb{E}\left[x\right]\right|^2\right]\right).
\end{aligned}
$$

Taking the first-order Taylor expansion of $\mathrm{U}(x)$ in the last three components of Equation (A.5) with respect to x, around $\mathbb{E}[x]$, provides[57]

$$
\begin{aligned}
II = &-\int_{x \le k}\left(\mathrm{U}\left(\mathbb{E}\left[x\right]\right) + \mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\left(x - \mathbb{E}\left[x\right]\right)\right)\frac{\Upsilon''\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}{\Upsilon'\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}\mathrm{E}\left[\varphi(x)\right]\mathrm{Var}\left[\varphi(x)\right]dx \\
&+\int_{x \ge k}\left(\mathrm{U}\left(\mathbb{E}\left[x\right]\right) + \mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\left(x - \mathbb{E}\left[x\right]\right)\right)\frac{\Upsilon''\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}{\Upsilon'\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}\mathrm{E}\left[\varphi(x)\right]\mathrm{Var}\left[\varphi(x)\right]dx \\
&+ R_{II,2}(x).
\end{aligned}
$$

Since $\mathbb{E}[x]$ is relatively close to k and $\mathrm{U}(k) = 0$, then $\mathrm{U}(\mathbb{E}[x]) \approx 0$. Therefore,

$$
II = \mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\int\left|x - \mathbb{E}\left[x\right]\right|\frac{\Upsilon''\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}{\Upsilon'\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}\mathrm{E}\left[\varphi(x)\right]\mathrm{Var}\left[\varphi(x)\right]dx + R_{II,2}(x).
$$

By Lemma 3, $\mathrm{Var}\left[\varphi(x)\right]$ is uncorrelated with $\left|x - \mathbb{E}\left[x\right]\right|\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}$. Therefore,

$$
\begin{aligned}
II = &\,\mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\int\mathrm{E}\left[\varphi(x)\right]\mathrm{Var}\left[\varphi(x)\right]dx\int\mathrm{E}\left[\varphi(x)\right]\left|x - \mathbb{E}\left[x\right]\right| \\
&\times\frac{\Upsilon''\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}{\Upsilon'\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}dx + R_{II,2}(x).
\end{aligned}
$$

Since $\left|x - \mathbb{E}\left[x\right]\right|$ is uncorrelated with $\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}$,

$$
\begin{aligned}
II = &\,\mathrm{U}'\left(\mathbb{E}\left[x\right]\right)\int\mathrm{E}\left[\varphi(x)\right]\mathrm{Var}\left[\varphi(x)\right]dx\int\mathrm{E}\left[\varphi(x)\right]\left|x - \mathbb{E}\left[x\right]\right|dx\int\mathrm{E}\left[\varphi(x)\right] \\
&\times\frac{\Upsilon''\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}{\Upsilon'\left(1 - \mathrm{E}\left[\mathrm{P}(x)\right]\right)}dx + R_{II,2}(x).
\end{aligned}
$$

Combining the LHS and the RHS (I and II), the uncertainty premium is then

$$
\mathcal{K} = -\frac{1}{2}\frac{\mathrm{U}''\left(\mathbb{E}[x]\right)}{\mathrm{U}'\left(\mathbb{E}[x]\right)}\mathbb{V}\mathrm{ar}\left[x\right] - \mathbb{E}\left[\frac{\Upsilon''(1-\mathrm{E}[\mathrm{P}(x)])}{\Upsilon'(1-\mathrm{E}[\mathrm{P}(x)])}\right]\mathbb{E}\left[\left|x - \mathbb{E}\left[x\right]\right|\right]\mho^2\left[x\right] + R_2(x).
$$

By I and II, $R_2(x) = o\left(\mathbb{E}\left[\left|x - \mathbb{E}\left[x\right]\right|^2\right]\right) + o\left(\int\mathrm{E}\left[\left|\varphi(x) - \mathrm{E}\left[\varphi(x)\right]\right|^3\right]x\,dx\right)$. Since, $o\left(\int\mathrm{E}\left[\left|\varphi(x) - \mathrm{E}\left[\varphi(x)\right]\right|^3\right]x\,dx\right)$ is equivalent to $o\left(\mathbb{E}\left[\left|x - \mathbb{E}\left[x\right]\right|^3\right]\right)$, then $R_2(x) = o\big(\mathbb{E}\big[\left|x - \mathbb{E}[x]\right|^2\big]\big)$ as $\left|x - \mathbb{E}\left[x\right]\right| \to 0$. □

[57] Note that II is of the order of cubic probabilities. Thus, it is smaller by two orders of magnitude than probabilities and therefore smaller by two orders of magnitude than I.

Proof of Corollary 1. Substituting for $\gamma = -\frac{\mathrm{U}''}{\mathrm{U}'}$ and $\eta = -\frac{\Upsilon''}{\Upsilon'}$ into Equation (11) of Theorem 7 completes the proof. □

Proof of Observation 1. Let the cardinality of $\mathcal{P}$ be $|\mathcal{P}| = n + 1$, where $n \geq 1$. The expected probability of $x \in X$ can then be written as

$$\mathrm{E}[\varphi(x)] = a_x + \frac{1}{n+1}\sum_{i=0}^{n}(b_x - a_x)\frac{i}{n} = a_x + \frac{1}{2}(b_x - a_x),$$

and the variance can be written as

$$\begin{aligned}\mathrm{Var}[\varphi(x)] &= \frac{1}{n+1}\sum_{i=0}^{n}\left(a_x + (b_x - a_x)\frac{i}{n} - \mathrm{E}[\varphi(x)]\right)^2 \\ &= \sum_{i=0}^{n}(b_x - a_x)^2\frac{i^2}{n^2(n+1)} - \frac{1}{4}(b_x - a_x)^2.\end{aligned}$$

The difference in $\mathrm{Var}[\varphi(x)]$ when $|\mathcal{P}| = n + 2$ is

$$\begin{aligned}D &= \sum_{i=0}^{n}(b_x - a_x)^2\frac{i^2}{n^2(n+1)} - \sum_{i=0}^{n+1}(b_x - a_x)^2\frac{i^2}{(n+1)^2(n+2)} \\ &= (b_x - a_x)^2\left(\sum_{i=0}^{n}\left(\frac{i^2}{n^2(n+1)} - \frac{i^2}{(n+1)^2(n+2)}\right) - \frac{1}{n+2}\right) \\ &= (b_x - a_x)^2\frac{1}{6n(n+1)}.\end{aligned}$$

This holds true for any $n \geq 1$. □

Proof of Observation 2. Clearly, $\mho^2 = 0$ only when $\mathcal{P}$ is a singleton. By Observation 1, the maximum variance of the probabilities of any $x \in X$ is attained when $|\mathcal{P}| = 2$, and each $\mathrm{P} \in \mathcal{P}$ assigns only the probabilities 0 or 1. In this case, $\mathrm{E}\left[\varphi_f(x)\right] = \xi$, and

$$\mathrm{Var}\left[\varphi_f(x)\right] = \xi(1-\xi)^2 + (1-\xi)(0-\xi)^2 = \xi - \xi^2,$$

which attains its maximum, $\mathrm{Var}[\varphi(x)] = \frac{1}{4}$, when $\xi = \frac{1}{2}$. When there are n outcomes with a nonzero probability, $\mathrm{E}\left[\varphi_f(x)\right] = \xi = \frac{1}{n}$, for each x. Therefore, the maximum value of $\mho^2$ is attained when there are only two outcomes with a nonzero probability. □

Proof of Observation 3. Clearly, $\mho^2 = 0$ only when $\mathcal{P}$ is a singleton. Suppose $\mathcal{P}$ consists of only the two equally likely dirac-delta probability distributions:

$$\delta_\alpha(x) = \begin{cases}\infty, & x = \alpha \\ 0, & x \neq \alpha\end{cases} \quad \text{and} \quad \delta_\beta(x) = \begin{cases}\infty, & x = \beta \\ 0, & x \neq \beta\end{cases},$$

where $\alpha \neq \beta$. Then, $\mathrm{E}[\varphi(\alpha)] = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}$, and $\mathrm{Var}[\varphi(\alpha)] = \lim_{\epsilon\to 0}\left(\frac{1}{2\epsilon}\right)^2$. The same holds true for β. Therefore, $\mho^2 = \lim_{\epsilon\to 0} 2\left(\frac{1}{2\epsilon}\right)\left(\frac{1}{2\epsilon}\right)^2 = \infty$. □

Proof of Observation 4. Since f and g induce the same partition on $\mathcal{S}$ into events,

$$\mho^2[f] = \int \mathrm{E}\left[\mathrm{P}\left(\{s \in \mathcal{S} \mid x - \epsilon < g(s) < x + \epsilon\}\right)\right] \times \mathrm{Var}\left[\mathrm{P}\left(\{s \in \mathcal{S} \mid x - \epsilon < g(s) < x + \epsilon\}\right)\right] dx = \mho^2[g],$$

where $\epsilon \to 0$. □

Proof of Observation 5. Immediately obtained from Observation 4. □

Proof of Observation 6. Let $f^* = h(f)$, where h is a strictly monotonic function. Because $\{s \in \mathcal{S} \mid x - \epsilon < f(s) < x + \epsilon\} = \{s \in \mathcal{S} \mid h(x) - \epsilon < f^*(s) < h(x) + \epsilon\}$ for every x, where $\epsilon \to 0$, the partition of $\mathcal{S}$ into events induced by f and by f^* is the same. Therefore, $\mho^2[f] = \mho^2\left[f^*\right]$. □

References

Anderson, E.W., Ghysels, E., Juergens, J.L., 2009. The impact of risk and uncertainty on expected returns. J. Financ. Econ. 94, 233–263.

Arrow, K.J., 1965. Aspects of the Theory of Risk Bearing. Yrjo Jahnssonin Saatio, Helsinki.

Augustin, P., Izhakian, Y., 2019. Ambiguity, volatility and credit risk. Rev. Financ. Stud. Forthcoming.

Bach, D.R., Hulme, O., Penny, W.D., Dolan, R.J., 2011. The known unknowns: neural representation of second-order uncertainty, and ambiguity. J. Neurosci. 31, 4811–4820.

Baillon, A., Bleichrodt, H., 2015. Testing ambiguity models through the measurement of probabilities for gains and losses. Am. Econ. J. Microecon. 7, 77–100.

Baillon, A., Huang, Z., Selim, A., Wakker, P.P., 2018. Measuring ambiguity attitudes for all (natural) events. Econometrica 86, 1839–1858.

Bayes, T., Price, R., Canton, J., 1763. An Essay Towards Solving a Problem in the Doctrine of Chances. C. Davis, Printer to the Royal Society of London, London, U.K.

Bernoulli, J., 1713. Ars Conjectandi (The Art of Conjecturing).

Bewley, T.F., 2011. Knightian decision theory and econometric inferences. J. Econ. Theory 146, 1134–1147.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81, 637–654.

Blackwell, D., Girshick, M., 1954. Theory of Games and Statistical Decisions. Wiley, New York.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. J. Polit. Econ. 96, 116–131.

Bollerslev, T., Sizova, N., Tauchen, G., 2011. Volatility in equilibrium: asymmetries and dynamic dependencies. Rev. Finance 16, 31–80.

Bossaerts, P., Ghirardato, P., Guarnaschelli, S., Zame, W.R., 2010. Ambiguity in asset markets: theory and experiment. Rev. Financ. Stud. 23, 1325–1359.

Boyle, P.P., Garlappi, L., Uppal, R., Wang, T., 2011. Keynes meets Markowitz: the tradeoff between familiarity and diversification. Manag. Sci. 58, 1–20.

Brenner, M., Izhakian, Y., 2018. Asset prices and ambiguity: empirical evidence. J. Financ. Econ. 130, 503–531.

Chen, Z., Epstein, L., 2002. Ambiguity, risk, and asset returns in continuous time. Econometrica 70, 1403–1443.

Chew, S.H., Sagi, J.S., 2008. Small worlds: modeling attitudes toward sources of uncertainty. J. Econ. Theory 139, 1–24.

Constantinides, G.M., 1982. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. J. Bus. 55, 253–267.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: local equity preference in domestic portfolios. J. Finance 54, 2045–2073.

Doan, B., Foster, F.D., Yang, L., 2018. Do investors differentiate between uncertainty and risk: evidence from trading around US macroeconomic news. Unpublished paper. Available on the "Auckland Center for Financial Research", https://acfr.aut.ac.nz/__data/assets/pdf_file/0006/186963/Do-investors-differentiate-between-uncertainty-and-risk_Bao-Doan.pdf.

Dow, J., Werlang, S., 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. Econometrica 60, 197–204.

Ellsberg, D., 1961. Risk, ambiguity, and the Savage axioms. Q. J. Econ. 75, 643–669.
Epstein, L.G., Ji, S., 2013. Ambiguous volatility and asset pricing in continuous time. Rev. Financ. Stud. 26, 1740–1786.
Epstein, L.G., Ji, S., 2014. Ambiguous volatility, possibility and utility in continuous time. J. Math. Econ. 50, 269–282.
Epstein, L.G., Schneider, M., 2007. Learning under ambiguity. Rev. Econ. Stud. 74, 1275–1303.
Epstein, L.G., Schneider, M., 2010. Ambiguity and asset markets. Annu. Rev. Financ. Econ. 2, 315–346.
Epstein, L.G., Zhang, J., 2001. Subjective probabilities on subjectively unambiguous events. Econometrica 69, 265–306.
Faria, G., Correia-da Silva, J., 2014. A closed-form solution for options with ambiguity about stochastic volatility. Rev. Deriv. Res. 17, 125–159.
Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J.F., Uribe, M., 2010. Risk matters: the real effects of volatility shocks. Am. Econ. Rev. 101, 2530–2561.
Ghirardato, P., Maccheroni, F., Marinacci, M., 2004. Differentiating ambiguity and ambiguity attitude. J. Econ. Theory 118, 133–173.
Ghirardato, P., Marinacci, M., 2002. Ambiguity made precise: a comparative foundation. J. Econ. Theory 102, 251–289.
Gilboa, I., Maccheroni, F., Marinacci, M., Schmeidler, D., 2010. Objective and subjective rationality in a multiple prior model. Econometrica 78, 755–770.
Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. J. Math. Econ. 18, 141–153.
Goetzmann, W.N., Kumar, A., 2008. Equity portfolio diversification. Rev. Finance 12, 433–463.
Goldberger, A., 1991. A Course in Econometrics, first ed.. Harvard University Press.
Grant, S., Karni, E., 2004. A theory of quantifiable beliefs. J. Math. Econ. 40, 515–546.
Halevy, Y., 2007. Ellsberg revisited: an experimental study. Econometrica 75, 503–536.
Hansen, L.P., Sargent, T.J., 2001. Robust control and model uncertainty. Am. Econ. Rev. 91, 60–66.
Hansen, L.P., Sargent, T.J., Tallarini, T.D., 1999. Robust permanent income and pricing. Rev. Econ. Stud. 66, 873–907.
Hao, L., Houser, D., 2012. Belief elicitation in the presence of naïve respondents: an experimental study. J. Risk Uncertain. 44, 161–180.
Hardy, G.H., Littlewood, J.E., Pólya, G., 1953. Inequalities, 2nd edition. Cambridge University Press, London. 1st edition, 1934.
Hayashi, T., Miao, J., 2011. Intertemporal substitution and recursive smooth ambiguity preferences. Theor. Econ. 6, 423–472.
Herron, R., Izhakian, Y., 2018. Ambiguity, risk, and dividend payout policy. SSRN eLibrary 2980600.
Hong, C.S., Wakker, P.P., 1996. The comonotonic sure-thing principle. J. Risk Uncertain. 12, 5–27.
Huettel, S.A., Stowe, C.J., Gordon, E.M., Warner, B.T., Platt, M.L., 2006. Neural signatures of economic preferences for risk and ambiguity. Neuron 49, 765–775.
Izhakian, Y., 2017. Expected utility with uncertain probabilities theory. J. Math. Econ. 69, 91–103.
Izhakian, Y., Benninga, S., 2011. The uncertainty premium in an ambiguous economy. Q. J. Financ. 1, 323–354.
Izhakian, Y., Yermack, D., 2017. Risk, ambiguity, and the exercise of employee stock options. J. Financ. Econ. 124, 65–85.
Izhakian, Y., Yermack, D., Zender, J., 2017. Ambiguity and the trade off theory of capital structure. SSRN eLibrary 2873248.
Izhakian, Y., Zender, J., 2018. Ambiguity and beliefs in the principal-agent model. SSRN eLibrary 3066761.
Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620.
Jewitt, I., Mukerji, S., 2017. Ordering ambiguous acts. J. Econ. Theory 171, 213–267.
Judd, K.L., 2003. Perturbation methods with nonlinear changes of variables. Mimeo, Hoover Institution.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.
Karlin, S., Taylor, H.M., 2012. A First Course in Stochastic Processes, second ed.. Elsevier Science.
Karni, E., Schmeidler, D., 1981. An expected utility theory for state-dependent preferences. Foerder Institute for Economic Research Working paper 48-80. Tel Aviv University.
Karni, E., Schmeidler, D., Vind, K., 1983. On state dependent preferences and subjective probabilities. Econometrica, 1021–1031.
Klibanoff, P., Marinacci, M., Mukerji, S., 2005. A smooth model of decision making under ambiguity. Econometrica 73, 1849–1892.
Knight, F.M., 1921. Risk, Uncertainty, and Profit. Houghton Mifflin, Boston.
Kopylov, I., 2007. Subjective probabilities on "small" domains. J. Econ. Theory 133, 236–265.
Kopylov, I., 2010. Simple axioms for countably additive subjective probability. J. Math. Econ. 46, 867–876.
Kothiyal, A., Spinu, V., Wakker, P.P., 2011. Prospect theory for continuous distributions: a preference foundation. J. Risk Uncertain. 42, 195–210.
Laplace, P.S., 1814. Théorie analytique des probabilitiés.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. 47, 13–37.
Maccheroni, F., Marinacci, M., Ruffino, D., 2013. Alpha as ambiguity: robust mean-variance portfolio analysis. Econometrica 81, 1075–1113.
Markowitz, H., 1959. Portfolio Selection: Efficient Diversification of Investments. Wiley, New York.
Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
Merton, R., 1976. Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. 3, 125–144.
Modigliani, F., Miller, M.H., 1958. The cost of capital, corporation finance and the theory of investment. Am. Econ. Rev. 48, 261–297.
Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica 34, 768–783.
Nau, R.F., 2006. Uncertainty aversion with second-order utilities and probabilities. Manag. Sci. 52, 136–145.
Papoulis, A., Pillai, S.U., 1965. Probability, Random Variables, and Stochastic Processes. Tata McGraw-Hill Education.
Pratt, J.W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122–136.
Rothschild, M., Stiglitz, J.E., 1970. Increasing risk: I. a definition. J. Econ. Theory 2, 225–243.
Rothschild, M., Stiglitz, J.E., 1972. Addendum to “increasing risk: I. a definition”. J. Econ. Theory 5, 306.
Rottenstreich, Y., Tversky, A., 1997. Unpacking, repacking, and anchoring: advances in support theory. Psychol. Rev. 104, 406–415.
Sarin, R.K., Wakker, P.P., 1992. A simple axiomatization of nonadditive expected utility. Econometrica 60, 1255–1272.
Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York, USA.
Schmeidler, D., 1989. Subjective probability and expected utility without additivity. Econometrica 57, 571–587.
Segal, U., 1987. The Ellsberg paradox and risk aversion: an anticipated utility approach. Int. Econ. Rev. 28, 175–202.
Seo, K., 2009. Ambiguity and second-order belief. Econometrica 77, 1575–1605.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finance 19, 425–442.
Shiller, R.J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? Am. Econ. Rev. 71, 421–436.
Siniscalchi, M., 2009. Vector expected utility and attitudes toward variation. Econometrica 77, 801–855.
Tobin, J., 1958. Liquidity preference as behavior towards risk. Rev. Econ. Stud. 25, 65–86.
Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323.
Tversky, A., Koehler, D.J., 1994. Support theory: a nonextensional representation of subjective probability. Psychol. Rev. 101, 547–567.
Ui, T., 2011. The ambiguity premium vs. the risk premium under limited market participation. Rev. Finance 15, 245–275.
Wakker, P.P., 1993. Unbounded utility for Savage’s “foundations of statistics,” and other models. Math. Oper. Res. 18, 446–485.
Wakker, P.P., 2010. Prospect Theory: For Risk and Ambiguity. Cambridge University Press, New York.
Wakker, P.P., Tversky, A., 1993. An axiomatization of cumulative prospect theory. J. Risk Uncertain. 7, 147–175.
Weil, P., 1989. The equity premium puzzle and the risk-free rate puzzle. J. Monet. Econ. 24, 401–421.