MODEL UNCERTAINTY

Massimo Marinacci
Department of Decision Sciences
and IGIER, Università Bocconi

Abstract
We study decision problems in which consequences of the various alternative actions depend on states determined by a generative mechanism representing some natural or social phenomenon. Model uncertainty arises because decision makers may not know this mechanism. Two types of uncertainty result, a state uncertainty within models and a model uncertainty across them. We discuss some two-stage static decision criteria proposed in the literature that address state uncertainty in the first stage and model uncertainty in the second (by considering subjective probabilities over models). We consider two approaches to the Ellsberg-type phenomena characteristic of such decision problems: a Bayesian approach based on the distinction between subjective attitudes toward the two kinds of uncertainty; and a non-Bayesian approach that permits multiple subjective probabilities. Several applications are used to illustrate concepts as they are introduced. (JEL: D81)

Kirk: *Do you think Harry Mudd is down there, Spock?*
Spock: *The probability of his presence on Motherlode is* 81% ± 0.53.

1. Introduction

In this section we briefly discuss several important notions—in particular, uncertainty (including model uncertainty), probabilities, and decisions. We then outline how the paper proceeds before making a few remarks on methodology.

The editor in charge of this paper was Juuso Valimaki.

Acknowledgments: This paper was delivered as the Schumpeter lecture at the 29th Congress of the European Economic Association in Toulouse in August 2014. It is dedicated, with admiration and affection, to David Schmeidler on the occasion of his 75th birthday. I thank Pierpaolo Battigalli, Loic Berger, Emanuele Borgonovo, Veronica Cappelli, Simone Cerreia-Vioglio, Fabio Maccheroni, Nicola Rosaia, Elia Sartori, and Juuso Valimaki as well as an anonymous referee for many comments and suggestions on preliminary versions of the paper. The material is based on extensive collaboration with Pierpaolo Battigalli, Simone Cerreia-Vioglio, Fabio Maccheroni, and Luigi Montrucchio; it is supported by the AXA Research Fund and the European Research Council (advanced grant BRSCDP-TEA).

E-mail: massimo.marinacci@unibocconi.it

Journal of the European Economic Association December 2015 13(6):998–1076
 DOI: 10.1111/jeea.12164

Uncertainty. Uncertainty has increasingly taken center stage in academic and public debates, and there is a growing awareness and concern about its role in human domains such as environmental uncertainty (climate change, natural hazards), demographic uncertainty (longevity and mortality risk), economic uncertainty (economic and financial crises), risk management (operational risks, Basel accords), and technological uncertainty (Fukushima).

Uncertainty affects decision making *directly* by making contingent the payoffs of a course of action (e.g., harvest and weather), as well as *indirectly* by generating private information. The latter point is key in strategic interactions, where uncertainty and private information are essentially two sides of the same coin: uncertainty generates private information when different agents have access to different information about the uncertain phenomenon; vice versa, private information per se can generate uncertainty if agents are contemplating it (moral hazard and adverse selection issues).

In the real world, uncertainty is a primary source of competitive advantage (and so of business opportunities that may favor entrepreneurship). In the theoretical world, uncertainty makes the study of agents' decisions and strategic interactions a beautiful and intellectually sophisticated exercise (altogether different from the study of physical particles' actions and interactions). In both worlds, uncertainty plays a major role.

Probabilities. Uncertainty and private information are thus twin notions. Uncertainty is indeed a form of partial knowledge (information) about the possible realizations of some contingencies that are relevant for agents' decisions (e.g., betting on a die: What face will come up?). As such, the nature of uncertainty is *epistemic*.[1] Intuitively, agents deal with uncertain contingencies by forming beliefs (expectations) about them. Yet how can the problem be properly framed? The notion of *probability* was the first key breakthrough: you can assign numbers to contingencies that quantify their relative likelihoods (and then manipulate those numbers according to the rules of probability calculus). Probability and its calculus emerged in the 16th and 17th centuries with the works of Cardano, Huygens, and Pascal, with a consolidation phase in the 18th and 19th centuries with the works of the Bernoullis, Gauss, and Laplace. In particular, the Laplace (1812) canon emerged, based on equally likely cases (alternatives): that the probability of an event is equal to the number of "favorable" cases divided by their total number.

Departing from the original epistemic stance of Laplace, over time the "equally likely" notion came to be viewed as a purely objective or physical feature (faces of a die, sides of a fair coin). Probability was no longer studied within decision problems, such as the games of chance that originally motivated its first studies in 16th and 17th centuries, but rather as a physical notion unrelated to decisions and therefore independent of any subjective information and beliefs. All this changed in the 1920s when de Finetti and Ramsey freed probability of physics,[2] put its study back in decision

1. From the Greek word for knowledge, episteme ($\varepsilon\pi\iota\sigma\tau\eta\mu\eta$). In the paper we use the terms "information" and "knowledge" interchangeably.

2. See Ramsey (1926) and de Finetti (1931, 1937).

problems (probability "is a measurement of belief qua basis for action", in Ramsey's words), and rendered "equally likely" an epistemic—and thus subjective—evaluation. They did so by identifying the probability that agents attach to some (decision relevant) event with their willingness to bet on it, which is a measurable quantity. As Ramsey remarked, the "old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds which he will accept".

Epistemic probabilities à la de Finetti–Ramsey (often called subjective) quantify decision makers' *degree of belief* and can be ascribed to any event, repeatable or not, such as "tomorrow it will rain" or "left-wing parties will increase their votes in the next elections". In this way, all uncertainty can be probabilized; this is the main tenet of Bayesianism.

Model Uncertainty. In this paper we consider decision makers (DMs) who are evaluating courses of actions the consequences of which depend on states of the environment that—such as rates of inflation, peak ground accelerations, and draws from urns—can be seen as realizations of underlying random variables that are part of a *generative* (or *data generating*) *mechanism* that represents some natural or social phenomenon.

Each such mechanism induces a probability model (or law) over states that describes the regular features of their variability. The uncertainty about the outcomes of the mechanism, and so about the inherent randomness of the phenomenon it represents, is called *physical*. Probability models thus quantify this kind of uncertainty, using analogies with canonical mechanisms (dice, urns, roulette wheels, and the like) that serve as benchmarks.[3] As any kind of uncertainty that DMs deem relevant for their decision problems, physical uncertainty is relative to their ex-ante (i.e., prior to the decision) information and its subjective elaboration—for instance, the analogy judgment just mentioned. Thus physical uncertainty (often referred to as risk in the literature) is an epistemic notion that accounts for DMs' views on the inherent randomness of phenomena. To paraphrase Protagoras: in decision problems, DMs are "the measure of all things".

Probability models describe such DMs' views by combining a structural component, which is based on theoretical knowledge (e.g., economic, physical), with a random component which accounts for measurement issues and for minor (and so omitted) explanatory variables.[4] We assume that DMs' ex-ante information allows them to posit a set of possible generative mechanisms, and so of possible probability models over states. Following a key tenet of classical statistics, we take such set as a datum of the decision problem. This set is generally nonsingleton (and so probabilities

3. In the words of Schmeidler (1989, p. 572), "the concept of objective probability is considered here as a physical concept like acceleration, momentum or temperature; to construct a lottery with given objective probabilities (a roulette lottery) is a technical problem conceptually not different from building a thermometer."

4. As Marschak (1953, p. 12) writes, they are "separately insignificant variables that we are unable and unwilling to specify". Similar remarks can be found in Koopmans (1947, p. 169); for a discussion, see Pratt and Schlaifer (1984, p. 12).

are "unknown") because the ex-ante information is not enough to pin down a single mechanism. Model uncertainty (or model ambiguity) therefore emerges since DMs are uncertain about the true mechanism.[5]

The often-made modeling assumption that a true generative mechanism exists is unverifiable in general and so of a metaphysical nature. It amounts to assuming that, among all probability models that DMs conceive, the model that best describes the variability in the states is the one that actually generates (and so causes/explains) them probabilistically. In any event, the assumption underlies a fruitful causal approach that facilitates the integration of empirical and theoretical methods—required for a genuine scientific understanding.[6]

Priors and Decisions. We assume that the DMs' ex-ante information also enables them to address model uncertainty through a subjective prior probability over models; in this we follow a key tenet of the Bayesian paradigm. Prior probabilities quantify DMs' beliefs by using analogies with betting behavior (Section 3.1).

The result is two layers of analysis: a first, classical layer featuring probability models on states that quantify physical uncertainty; and a second, Bayesian layer characterized by a prior probability on models that quantifies model uncertainty. As is well known, both layers involve nontrivial methodological aspects. The second layer is ignored by classical statistics; the first layer is indirectly considered within the Bayesian paradigm through arguments of de Finetti representation theorem type.[7]

However, our motivation is pragmatic: we expect that the uncertainty characterizing many decision problems that arise in applications can be fruitfully analyzed by distinguishing physical and model uncertainty within DMs' ex-ante information.[8]

5. See Wald (1950), Fisher (1957), Neyman (1957), and Haavelmo (1944, pp. 48–49). We will use "model uncertainty" throughout even though "model ambiguity" is a more specific and hence more informative terminology (see Hansen 2014). In any case, at the level of generality of our analysis, model uncertainty is an all-encompassing notion. We abstract from any finer distinction, say between nonparametric (model) and parametric (estimation) uncertainty—that is, between models that differ either in substance (e.g., Keynesian or New Classical specifications in monetary economics) or in detail (e.g., different coefficient values within a theoretical model). See Hansen (2014, p. 974) and Hansen and Sargent (2014, p. 1) for a related point. For finer distinctions, see for example Draper et al. (1987), Draper (1995), the references therein (these authors call predictive uncertainty a notion similar to physical uncertainty), as well as Brock, Durlauf, and West (2003).

6. For a critique of this approach, with a purely descriptive, acausal, interpretation of models, see Akaike (1985), Breiman (2001), and Rissanen (2007) as well as Barron, Rissanen, and Yu (1998), Hansen and Yu (2001), and Konoshi and Kitagawa (2008). Descriptive approaches now play an important role in information technology thanks to the large data sources currently available, the so-called big data (see, e.g., Halevy, Norvig, and Pereira 2009).

7. Cerreia-Vioglio et al. (2013a) provide a decision-theoretic derivation of the two layers within a de Finetti perspective. However, such asymptotic perspective (based on exchangeability and related large-sample properties) may be a straightjacket when DMs have enough information to directly specify a set of probability models. It is typical in these cases for models to be identified by parameters that have some concrete meaning beyond their role as indices (urns' compositions being a prototypical example). See Section 3.2.

8. As Winkler (1996) notes, this distinction can be seen as an instance of the "divide et impera" precept, a most pragmatic principle. A related point is eloquently made by Ekeland (1993, pp. 139–146). It is

The two layers of analysis motivated by such a distinction naturally lead to two-stage decision criteria: actions are first evaluated with respect to each possible probability model, and then such evaluations are combined by means of the prior distribution. In other words, in this hierarchical approach we first assess actions in terms of physical uncertainty and then in terms of epistemic uncertainty.[9]

Outline. Our paper is an overview of both traditional and more recent elaborations of this basic insight. In fact, the failure to distinguish these two kinds of uncertainty—in particular, the specific role of model uncertainty—may have significant economic consequences. Both behavioral paradoxes (of which Ellsberg's is the most well known) and empirical puzzles (e.g., in asset pricing) can be seen as the outcome of a too limited account of uncertainty in standard economic modeling. To emphasize their relevance, we will illustrate concepts with several applications; these include structural reliability (Section 4.5), monetary policy (Section 4.9), asset pricing (Section 4.10), and public policy (Section 4.11).[10] We begin in Section 2 by presenting classical decision problems, the two-stage static decision framework of the paper; a few examples from different fields are given to illustrate its scope. In Section 3 we introduce a two-stage expected utility criterion that reduces epistemic uncertainty to physical uncertainty ("uncertainty to risk" as it is often put), and so ignores the distinction. Yet given that experimental and empirical evidence indicate that this distinction is relevant and may affect valuation, in Sections 4 and 5 the criterion is modified in two different ways, Bayesian (Section 4) and not (Section 5), in order to deal more properly with model uncertainty. In these sections we discuss how optimal behavior is affected by model uncertainty as well as the extent to which such behavior reflects a desire for robustness that, in turn, may favor action diversification, lead to no trade results, and result in market prices for assets that—by incorporating premia for model uncertainty—may explain some empirical puzzles (Sections 4.9–4.11 and 5.3).

A final important remark: although some important results have already been established for dynamic choice models under ambiguity, we consider only static choice models.[11]

Methodological Post Scriptum. In a philosophy of probability perspective, the two layers of our analysis rely on different meanings of the notion of probability. In

interesting that de Finetti (1971, p. 89) acknowledge that "... it may be convenient to use the 'probability of a law (or theory)' as a useful mental intermediary to evaluate the probability of some *fact* of interest" (emphasis in the original).

9. With all the relevant caveats, for brevity we will often refer to model uncertainty as epistemic uncertainty, which is short for "epistemic uncertainty about models" (and where by models we mean probability distributions over states).

10. Though our focus is economics, we discuss applications in other disciplines to place concepts in perspective.

11. We refer to Miao (2014) for a recent textbook exposition; for different perspectives on the topic and on the literature (see Hanany and Klibanoff 2009; Siniscalchi 2011; Strzalecki 2013).

particular, a distinction is often made, mostly after Carnap (1945, 1950),[12] between epistemic and physical uncertainties, which are quantified (respectively) by epistemic and physical probabilities.[13] In this dual view, model uncertainty is regarded as a main example of epistemic uncertainty, with prior probabilities representing DMs' degrees of beliefs about models.[14] Yet for decision problems this distinction is questionable at an ontological level because, as mentioned previously, all uncertainty that is relevant to decision problems is, ultimately, epistemic (i.e., relative to the state of DMs' information). In this paper we regard physical probability as the DMs modeling of the variability in the states of the environment, which they can carry out through analogies—"as if" arguments—with canonical mechanisms (e.g., states may be viewed to obtain *as if* they were colors drawn from urns).[15] For our probabilistic understanding is shaped, nolens volens, by such mechanisms, and it was their role in games of chance that actually gave rise to probability in the 16th and 17th centuries.[16] In turn, probabilities in canonical mechanisms—and then, by analogy, in general settings—can be interpreted as physical concepts in potential terms (dispositions, say à la Popper

12. See Good (1959, 1965), Hacking (1975), Shafer (1978), von Plato (1994), and Cox (2006) for discussions on this distinction. More applied perspectives can be found in Apostolakis (1990), Paté-Cornell (1996), Walker et al. (2003), Ang and Tang (2007), Der Kiureghian and Ditlevsen (2009), and Marzocchi, Newhall, and Woo (2012) as well as in the 2009 report of the National Research Council. This distinction between the two notions of probability traces back to Cournot and Poisson, around the year 1840 (see, e.g., Zabell 2011); Keynes (1921, p. 312) credits Hume for an early dual view of probability. LeRoy and Singell (1987) discuss a related distinction made in Knight (1921).

13. Terminology varies: in place of "physical" the terms *aleatory* and *objective* are often used, as are the terms *phenomenological* (Cox 2006) and *statistical* (Carnap 1945); in place of "epistemic" the term *subjective* is often used, as is (though less frequently) *personal* (Savage 1954). Finally, physical probabilities are sometimes called *chances*, with the term *probability* being reserved for epistemic probabilities (Anscombe and Aumann 1963; Singpurwalla 2006; Lindley 2013).

14. In the logical approaches of Carnap and Keynes, they actually represent degrees of confirmation (see, e.g., Carnap 1945, p. 517). The subjectivist approach, most forcefully proposed by de Finetti, is today more widely adopted.

15. Le Cam (1977, p. 154) writes ". . . most models or theories of nature which are encountered in statistical practice are probabilistic or stochastic. The probability measures entering in these models are . . . used to indicate a certain structure which can, in final analysis, be reduced to this 'Everything is as if one were drawing balls from a well-mixed bag.'" Much earlier, Borel (1909, p. 167) had written "we can now for formulate the general problem of mathematical statistics as follows: determine a system of drawing made of urns having a fixed composition, so that the results of a series of drawings, interpreted with the aid of coefficients conveniently selected, lead with a great likelihood to a table which is identical with the table of observations" (p. 138 of the 1965 English translation). More recently, Gilboa, Lieberman, and Schmeidler (2010) propose a view of probability based on a formal notion of similarity.

16. Hacking (1975) is a well-known account of early probability thinking (for a discussion, see Gilboa and Marinacci 2013). On infants and urns, see Xu and Garcia (2008) and Xu and Kushnir (2013); possible neurological bases of probabilistic reasoning are discussed by Kording (2007), and its importance, as a part of unconscious cognitive abilities, is discussed by Tenenbaum et al. (2011). They contrast conscious and unconscious manipulations of probabilities and explain the former's problems by noting that numerical probabilities are "a recent cultural invention that few people become fluent with, and only then after sophisticated training". These remarks are reminiscent of the expert billiard player example famously used by Friedman and Savage (1948) and Friedman (1953) to illustrate the "as if" methodology—which, as they write on p. 298, ". . . does not assert that individuals explicitly or consciously calculate and compare expected utilities . . . but behave *as if* they . . . [did] . . . " (emphasis in the original).

1959) or in actual terms (frequencies, say à la von Mises 1939).[17] That being said, in stationary environments these two interpretations can be reconciled via ergodic arguments.[18]

2. Setup

2.1. Notation and Terminology

Probability Measures. Let (S, Σ) be a measurable space, where Σ is an algebra of events of S (events are always understood to be in Σ). For instance, Σ could be the power set 2^S of S—that is, the collection of all subsets of S. In particular, when S is finite we assume that $\Sigma = 2^S$ unless otherwise stated.

Let $\Delta(S)$ be the collection of all (countably additive) probability measures $m : \Sigma \to [0, 1]$. If S is a finite set with n elements and Σ is the power set of S, then $\Delta(S)$ can be identified with the simplex $\Delta_{n-1} = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ of $\mathbb{R}^n$.

We will consider probability measures μ defined on the power set of $\Delta(S)$; for simplicity, we assume that they have finite nonempty support. In other words, we assume that there exists a finite subset of $\Delta(S)$, denoted by $\operatorname{supp}\mu$, such that $\mu(m) > 0$ if and only if $m \in \operatorname{supp}\mu$. Given any subset $M \subseteq \Delta(S)$, we denote by $\Delta(M)$ the collection of all probability measures μ with $\operatorname{supp}\mu \subseteq M$.

Integrals and Sums. Given a measurable space $(X, \mathcal{X})$, we often use the shorthand notation $\mathrm{E}_p f$ to denote the (Lebesgue) integral $\int_X f(x)dp(x)$ of a $\mathcal{X}$-measurable function $f : X \to \mathbb{R}$ with respect to a probability measure $p : \mathcal{X} \to [0, 1]$. In particular, if the space X itself is finite, then

$$\int_X f(x)dp(x) = \sum_{x \in X} f(x)p(x)$$

17. Popper (1959, p. 37) mentions these Aristotelian categories, though he claims that "propensities . . . cannot . . . be inherent in the die, or in the penny, but in something a little more abstract, even though physically real: they are relations properties of the experimental arrangement—of the conditions we intend to keep constant during repetition." Earlier in the paper, on p. 34, he writes that "The frequency interpretation always takes probability as relative to a sequence . . . is a *property of some given sequence*. But with our modification, the sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be a *property of the generating conditions*" (emphasis in the original).

18. In the terminology of Section 3.2, let p be a probability on Σ. Given any event $B_k \subseteq \times_{i=1}^k Z$, define the empirical measure by $\hat{p}_T(B_k)(s) = (1/T)\sum_{i=1}^T 1_{((\tilde{z}_i,\ldots,\tilde{z}_{i+k})\in B_k)}(s)$. If process $\{\tilde{z}_t\}$ is stationary and ergodic, a well-known consequence of the individual ergodic theorem is that, p-almost everywhere, $\lim_T \hat{p}_T(B_k) = p(B_k)$. So in this case the probability p can be interpreted as the limit empirical frequency with which states occur. Von Plato (1988, 1989) emphasize this "time average" view of frequentism, which can be reconciled with propensities when $p(B_k)$ is interpreted as a propensity (a similar remark is made in an i.i.d. setting by Giere 1975, p. 219).

that is, integrals reduce to sums. Throughout the paper, the reader can always assume that spaces are finite and so integrals can be interpreted as sums (an exception are the examples that involve normal distributions, which require infinite spaces). In this regard, note that sums actually arise even in infinite spaces when the support of the probability is finite (and measurable)—that is, supp $p = \{x \in X : p(x) > 0\}$. In this case, $\int_X f(x)dp(x) = \sum_{x \in \text{supp}\, p} f(x)p(x)$.

Differentiability. The presentation will require us to consider differentiability on sets that are not necessarily open. To ease matters, throughout the paper we say that a (convex) function $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ is *differentiable* on a (convex) set C if it can be extended to a (convex) differentiable function on some open (convex) set containing C. If the set C is open, such set is C itself.

Equivalent and Orthogonal Measures. Two probability measures m and $\tilde{m}$ in Δ are *orthogonal*, written $m \perp \tilde{m}$, if there exists $E \in \Sigma$ such that $m(E) = 0 = \tilde{m}(E^c)$; here E^c denotes the complement of E. In words, two orthogonal probabilities assign zero probability to complementary events. A finite collection of measures $M = \{m_1, \dots, m_n\} \subseteq \Delta$ is *(pairwise) orthogonal* if all its elements are pairwise orthogonal, that is, there exists a measurable partition $\{E_i\}_{i=1}^n$ of events such that, for each i, $m_i(E_i) = 1$ and $m_i(E_j) = 0$ if $j \neq i$.

We say that m is *absolutely continuous* with respect to $\tilde{m}$, written $m \ll \tilde{m}$, if $\tilde{m}(E) = 0$ implies $m(E) = 0$ for all events E. The two measures are *equivalent* if they are mutually absolutely continuous, that is, if they assign zero probability to the same events.

2.2. Decision Form

Following Wald (1950), a decision problem under uncertainty consists of a *decision maker* (DM) who must choose among a set of alternative actions whose consequences depend on uncertain factors that are beyond his control. Formally, there is a set A of available *actions* a that can result in different material *consequences* c, within a set C, depending on which *state of the environment* s obtains in a state space S. As discussed in the Introduction, states are viewed as realizations of some underlying random variables. Often we consider monetary consequences; in such cases, C is assumed to be an interval of the real line.

The dependence of consequences on actions and states is described by a *consequence function* $\rho : A \times S \to C$ that details the consequence

$$c = \rho(a, s) \tag{1}$$

of action a in state s. The quartet (A, S, C, ρ) is a *decision form under uncertainty*. It is a static problem that consists of an ex-ante stage (up to the time of decision) and an ex-post stage (after the decision). Ex ante, DMs know all elements of the quartet (A, S, C, ρ) and, ex post, they will observe the consequence $\rho(a, s)$ that results.

However, we do not assume that DMs will necessarily observe the state that, ex post, obtains.[19]

We illustrate decision forms with a few examples from different fields. Some of them will also be used later in the paper to illustrate various concepts as they are introduced.

Betting. Gamblers have to decide which bets to make on the colors of balls drawn from a given urn. The consequence function is $\rho(a,s) = w(s,a) - c(a)$, where $w(a,s)$ is the amount of money that bet a pays if color s is drawn and $c(a)$ is the price of the bet. States (i.e., balls' colors) are observed ex post.

Monetary Policy. Monetary authorities have to decide on the target level of inflation that will best control the economy's unemployment and inflation (Sargent 2008). More specifically, consider a class $\theta \in \Theta$ of linear model economies in which unemployment and inflation outcomes (u, π) are related to shocks (w, ε) and to the government policy a as follows:

$$\begin{aligned} u &= \theta_0 + \theta_{1\pi}\pi + \theta_{1a}a + \theta_2 w \\ \pi &= a + \theta_3\varepsilon. \end{aligned}$$

The vector parameter $\theta = (\theta_0, \theta_{1\pi}, \theta_{1a}, \theta_2, \theta_3) \in \mathbb{R}^5$ specifies the relevant structural coefficients. Coefficients $\theta_{1\pi}$ and θ_{1a} are slope responses of unemployment to actual and planned inflation, while the coefficients θ_2 and θ_3 quantify shock volatilities. Finally, the intercept θ_0 is the rate of unemployment that would (systematically) prevail in the absence of policy interventions.[20]

States have random and structural components $s = (w, \varepsilon, \theta) \in W \times E \times \Theta$. Consequences are the unemployment and inflation pairs $c = (u, \pi) \in \mathbb{R}_+ \times \mathbb{R}$. Since the reduced form of each model economy is

$$\begin{aligned} u &= \theta_0 + (\theta_{1\pi} + \theta_{1a})a + \theta_{1\pi}\theta_3\varepsilon + \theta_2 w \\ \pi &= a + \theta_3\varepsilon \end{aligned}$$

it follows that the consequence function has the form

$$\rho(a, w, \varepsilon, \theta) = \begin{bmatrix} \theta_0 \\ 0 \end{bmatrix} + a \begin{bmatrix} \theta_{1\pi} + \theta_{1a} \\ 1 \end{bmatrix} + \begin{bmatrix} \theta_2 & \theta_{1\pi}\theta_3 \\ 0 & \theta_3 \end{bmatrix} \begin{bmatrix} w \\ \varepsilon \end{bmatrix}. \tag{2}$$

19. If DMs do observe the state then, in a temporal setting, both consequences and states (past and current) may become data available for future decisions. In order to address that possibility we augment the problem's structure with a feedback function that specifies what is observed ex post. In dynamic setups—where today's ex post is tomorrow's ex ante—feedback plays a key role (see Battigalli et al. 2015a).

20. See Battigalli et al. (2015b), who interpret this atemporal setup as a stochastic steady state.

The policy multiplier is $\theta_{1\pi} + \theta_{1a}$. For instance, a zero multiplier (i.e., $\theta_{1a} = -\theta_{1\pi}$) characterizes a Lucas–Sargent model economy in which monetary policies are ineffective, whereas $\theta_{1a} = 0$ characterizes a Samuelson–Solow model economy in which such policies may be effective.

Finally, states—that is, shocks' realizations and the true model economy—are not necessarily observed ex post.

Production. Firms have to decide on the level of production for some output even when they are uncertain about the price that will prevail. The consequence function is the profit $\rho(a, s) = r(s, a) - c(a)$, where $r(s, a)$ is the revenue generated by a units of output under price s and $c(a)$ is the cost of producing a units of output. In this case, states—that is, the price of output—are observed ex post.

Inventory. Retailers have to decide which quantity of some product to buy wholesale, while uncertain about how much they can sell.[21] A retailer can buy any quantity a of the product at a cost $c(a)$. An unknown quantity s of the product will be demanded at a unit price p. Here the consequence function is the profit $\rho(a, s) = p \min\{a, s\} - c(a)$, where $\min\{a, s\}$ is the amount that retailers will actually be able to sell. States—the demand for the product—are observed ex post.

Financial Investments. Investors have to decide how to allocate their wealth among some financial assets traded in a frictionless financial market. Suppose there are n such assets at the time of the decision, each featuring an uncertain gross return r_i after one period. Denote by $a \in \Delta_{n-1}$ the vector of portfolio weights, where a_i indicates the fraction of wealth invested in asset $i = 1, \ldots, n$. Given an initial wealth w, the consequence function $\rho(a, s) = (a \cdot s)w$ is the end-of-period wealth determined by a choice a when the vector $s = (r_1, \ldots, r_n) \in \mathbb{R}^n$ of returns obtains. States—here, the returns on assets—are again observed ex post.

Climate Change Mitigation. Environmental policy makers have to decide on the abatement level of gas emissions in attempting to mitigate climate change (see, e.g., Gollier 2013; Millner, Dietz, and Heal 2013; Berger, Emmerling, and Tavoni 2014). The state space S consists of the possible climate states, which may consist of structural and random components about the cause–effect chains from emissions to temperatures (Meinshausen et al. 2009). For instance, states may be represented by (equilibrium) climate sensitivity, a quantity that measures the equilibrium global average surface warming that follows a doubling of atmospheric carbon dioxide (CO_2) concentrations (Solomon et al. 2007).

The action space A consists of all possible abatement level policies. The consequence function $\rho(a, s) = d(s, a) - c(a)$ describes the overall monetary consequence of abatement policy a when s is the climate state, as determined by the monetary damage $d(s, a)$ and by the abatement cost $c(a)$.

21. This decision problem is often called the *newsvendor* problem (see, e.g., Porteus 2002).

Natural Hazards. Civic officials have to decide whether or not to evacuate an area because of a possible natural hazard (see, e.g., Marzocchi, Newhall, and Woo 2012). In the case of earthquakes, for instance, the state space S may consist of all possible peak ground accelerations (PGAs) that describe ground motion; the action space A consists of the two actions a_0 (no evacuation) and a_1 (evacuation), and the consequence function $\rho(a, s)$ describes the consequence (in monetary terms) of action a when s is the PGA that obtains.[22] We distinguish different components in such consequence: (i) the damage to infrastructures $d_b(s)$ and the human casualties $d_h(s)$ that a PGA s determines, (ii) the evacuation cost δ.[23] Since evacuation can, essentially, only reduce the number of human casualties, we can write the function $\rho : A \times S \to \mathbb{R}$ as follows:

$$\rho(a, s) = \begin{cases} d_b(s) + d_h(s) & \text{if } a = a_0, \\ d_b(s) + \delta & \text{if } a = a_1. \end{cases}$$

As to the damage functions $d_b : S \to \mathbb{R}$ and $d_h : S \to \mathbb{R}$, note that:

- their argument s is a physical magnitude, the PGA, related to a Richter-type scale;
- their images $d_b(s)$ and $d_h(s)$ are a socio-economic magnitude, related to a Mercalli-type scale;
- their graphs $\mathrm{Gr}\, d_b = \{(s, c) : c = d_b(s)\}$ and $\mathrm{Gr}\, d_h = \{(s, c) : c = d_h(s)\}$ represent both aspects.

Finally, states—the PGAs—are observed ex post.

Structural Reliability. Structural engineers have to decide what design of a structure will make it most reliable (Ditlevsen and Madsen 2007). Consider for example the design of a cantilever beam of some fixed length (Field and Grigoriu 2007). The action—in other words, the design variable that the engineer must choose—is the square cross-section of that beam. The stiffness s of the beam, which is unknown because of the physical uncertainty affecting material properties (Guo and Du 2007, p. 2337), determines the tip deflection $\tau(a, s)$ of the beam resulting from the choice of the square cross-section a. The beam breaks (and so a structural failure occurs) if $\tau(a, s) > d$, where $d \geq 0$ is the maximum tip displacement. Hence $F_a = \{s \in S : \tau(a, s) > d\}$ and $F_a^c = \{s \in S : \tau(a, s) \leq d\}$ are, respectively, the failure event and the safe event determined by action a. The consequence function

22. If $s = 0$, no ground motion occurred. The civic officials may rely on seismologists to identify the relevant state space, and to structural engineers and economists to assess the consequence function (in this evacuation decision structures are a given; a design decision problem will be considered next).

23. In Marzocchi, Newhall, and Woo (2012), C corresponds to $-\delta$ and L to a constant damage function d_h—that is, $d_h(s) = -L$ for each s. In their words "the principal protection cost C is the economic dislocation which may last for weeks or even months" and "for a non-evacuation decision, the principal loss incurred is that of human life [. . .] usually measured using the economic concept of Willingness to Pay for Life Saved". On d_h, see also Porter, Shoaf, and Seligson (2006).

is given by

$$\rho(a,s) = \begin{cases} \delta + c(a) & \text{if } s \in F_a, \\ c(a) & \text{else,} \end{cases}$$

where δ is the damage cost of failure and $c(a)$ is the cost of the square cross-section a. States—the beam's stiffness—might not be observed ex post.

Quality Control. Managers have to decide whether to accept or reject the shipment of some parts from a supplier—say, integrated circuits from an electrical company (see, e.g., Raiffa and Schlaifer 1961; Berger 1993). The state space S consists of the proportion s of defective circuits in the shipment. Only the whole shipment can be rejected, not individual parts. Thus the action space A consists of the two actions, a_0 (reject the shipment) and a_1 (accept the shipment). The consequence function is given by $\rho(a_0, s) = 0$ and $\rho(a_1, s) = r(s) - c(s) - p$; here $r(s)$ is the revenue from the sales of the output produced when s is the proportion of defective circuits, $c(s)$ is the cost they entail when entering the production line (Deming 1986, Chap. 15), and p is the price of the shipment once accepted. In other words: if the shipment is rejected, then there is no output and so no revenues; if the shipment is accepted, the consequence depends on its cost and on the revenues and costs that it determines given the proportion of defective circuits that it features. States—here, the proportion of defective circuits—are observed ex post unless defects are latent (Deming 1986, p. 409).

Public Policy. Public officials have to decide which treatment—for example, which type of vaccination—should be administered to individuals who belong to a heterogeneous population that, for policy purposes, is classified in terms of some observable characteristic (covariate), such as age or gender.

Let X and T be, respectively, the (finite) collections of covariates and of possible alternative treatments. If only aggregate (and not individual) outcomes matter for policy making, then we can regard actions as functions $a : X \to \Delta(T)$ that associate probability distributions over treatments with covariates.[24] Here $a(x)(t) \in [0, 1]$ is the fraction of the population with covariate x that has been assigned treatment t.[25] Treatment actions are *fractional* if they do not assign the same treatment to all individuals with the same covariate (see Manski 2009).

When state s obtains, $c_x(t, s)$ denotes the (scalar) outcome for individuals with covariate x who have been assigned treatment t.[26] If public officials care about the

24. This simple distributional formulation abstract from the individual choice problems that may affect treatment effects, which is a key issue in actual policy analysis (see Heckman 2008, p. 7).

25. The treatment allocation can be implemented through an anonymous random mechanism (a form of the "equal treatment of equals" ethical principle; see Manski 2009). If the population is large, then under standard assumptions $a(x)(t)$ can be regarded as the fraction of the population with covariate x under treatment t and also as the probability with which the random mechanism assigns the treatment to every individual with that covariate.

26. The outcome can be material (say, monetary) or, from a utilitarian perspective, can be stated in terms of welfare (say, in utils). Our specification presumes that the relevant policy information is about

average of such individual outcomes, then the consequence function is $\rho(a,s) = \sum_x (\sum_{t\in T} c_x(t,s)a(x)(t))p(x)$, where $p(x)$ is the fraction of the population with covariate x. The existence of various states explains why individuals with the same covariate may respond differently to the same treatment.

Epistemology. Scientists have to decide whether to adopt a new theory or retain an old one. We can follow Giere (1985, pp. 351–352) and consider Earth scientists who in the 1960s were faced with deciding whether to accept the new drift hypotheses or retain the old static ones. Abstracting from peer concerns, there are two relevant states: "the tectonic structure of the world is more similar to drift models than to static models" and the reverse, as well as two actions, "adopt drift models" and "retain static models". As Giere argues, status quo biases determine the consequence function that characterizes the decision problem, which can be described by the following table:

	Drift models approximately correct	Static models approximately correct
Adopt	Satisfactory	Terrible
Retain	Bad	Excellent

Interactive Situations. Players have to decide which actions to play in a static interactive situation. For instance, in the previous production example suppose there is a set $I = \{1,\ldots,n\}$ of firms (an oligopoly) and that the price depends on their aggregate production $\sum_{i\in I} a_i$ according to a (commonly known) inverse demand function $D^{-1}(\sum_{i\in I} a_i)$. In the decision problem of each firm i, the state is no longer the price but rather the production profile $a_{-i} = (a_j)_{j\neq i}$ of the other firms; that is, $S = A_{-i} \equiv \times_{j\neq i} A_j$. In fact, for firm i that profile determines the monetary consequence $\rho_i(a_i, a_{-i}) = r_i(a_i, D^{-1}(\sum_{i\in I} a_i)) - c_i(a_i)$. The firm's decision problem is thus $(A_i, A_{-i}, C_i, \rho_i)$, where $C_i \subseteq \mathbb{R}$.

In general, a static game form $G = (I, (C_i)_{i\in I}, (A_i)_{i\in I}, (g_i)_{i\in I})$ among selfish players[27] consists of a set $I = \{1,\ldots,n\}$ of players and, for each player i, a set C_i of consequences, a set A_i of actions, and an outcome function $g_i : A_1 \times \cdots \times A_n \to C_i$ that associates the material outcome $g_i(a_1,\ldots,a_n)$ with action profile $(a_1,\ldots,a_n)$.

When player $i \in I$ evaluates action a_i, the relevant states s are the action profiles a_{-i} of his opponents. As a result, the state space $S = A_{-i}$ is the collection of all action profiles of his opponents and the consequence function is his individual outcome function $g_i : A_i \times A_{-i} \to C_i$. The decision form for player i is thus (A_i, A_{-i}, C_i, g_i). States (i.e., opponents' actions) may be observed ex post.

individuals with given covariates (say, the effect of vaccination on elderly white males) and not about particular individuals.

27. That is to say, players who (like the firms just described) care only about their own material outcomes. In this paper we do not consider other regarding preferences (for a generalization of standard preferences to such case, see Maccheroni, Marinacci, and Rustichini 2012).

2.3. Decision Problem

Decision making is the outcome of DMs' mental elaboration of their desires and beliefs. We thus assume that DMs have a preference $\succsim$ over actions, a binary relation that describes how they rank alternatives. In particular, we write $a \succsim b$ if the DM either strictly prefers action a to action b or is indifferent between the two. As usual, $\sim$ denotes indifference and $\succ$ strict preference. The quintet $(A, S, C, \rho, \succsim)$ is a *decision problem under uncertainty*. The aim of DMs is to select actions that are *optimal* with respect to their preference—that is, actions $\hat{a} \in A$ such that $\hat{a} \succsim a$ for all actions $a \in A$.[28]

For each action $a \in A$, the *section* $\rho_a : S \to C$ of ρ at a defined by

$$\rho_a(s) = \rho(a, s)$$

associates with each $s \in S$ the consequence resulting from the choice of a if s obtains. From the consequentialist perspective, what matters about actions is not their label/name but instead the consequences that they determine under different states. This viewpoint motivates the following classical principle.

CONSEQUENTIALISM. Two actions that are realization equivalent (i.e., that generate the same consequence in every state) are indifferent. Formally,

$$\rho(a, s) = \rho(b, s) \quad \forall s \in S \Longrightarrow a \sim b$$

or, equivalently, $\rho_a = \rho_b \Longrightarrow a \sim b$.

Consequentialism is trivially satisfied when the consequence function is such that $a = b$ whenever $\rho(a, s) = \rho(b, s)$ for each $s \in S$. For example, this is the case for the consequence function (2) of the monetary policy example.

Assume that, given any consequence c, there is a "sure" action a_c that delivers this consequence in all states; that is, $\rho(a_c, s) = c$ for all $s \in S$. In that case, we can define a derived preference $\succsim_C$ among consequences as $c \succsim_C c'$ if and only if $a_c \succsim a_{c'}$. By Consequentialism, this definition is well posed. To ease notation, we will more simply write $c \succsim c'$.[29]

28. Preferences are here viewed as mental constructs, with a cognitive appeal and meaning (Skinner 1985, pp. 295–296, sketches a behaviorism interpretation of some basic decision theory notions, without cognitive notions). Though disciplined by them (the more, the better, obviously), they are not just ways to organize behavioral data. Preferences thus have here a more substantive interpretation than the one envisioned by Pareto (1900), in the paper that started the ordinalist revolution, and often adopted by the revealed preference literature that followed the seminal (1938) work of Samuelson. According to Pareto, p. 222, "If a dog . . . leaves the soup and takes the meat . . . [this behavior] can be expressed in the phrase that this dog *prefers* the meat to the soup . . . But the intention is to express only the fact observed: the word *prefer* should not be taken as implying any judgement on the part of the animal or any comparison between two kinds of pleasure" (emphasis in the original).

29. Although the preference between consequences has been derived from the primitive preference among actions, at a conceptual level the opposite might well be true: DMs may actually have "basic preferences" among consequences, which in turn determine how they rank actions. Yet actions are the objects of choice

2.4. Savage Acts

Under Consequentialism, we can define a preference on sections as follows:

$$\rho_a \succsim \rho_b \Longleftrightarrow a \succsim b \tag{3}$$

That is, the ranking on actions can be translated into a ranking of sections. This possibility suggests a more radical approach, adopted by Savage (1954), whereby actions are identified if they are realization equivalent. Thus, any two actions that deliver the same consequences in the various states are identified—no matter how different such actions might be in other regards. For instance, if different production levels result in the same profits in all states, we identify them: they can be regarded as equivalent from a decision-theoretic standpoint.[30]

In most of what follows we adopt Savage's approach. As a result, in place of actions we consider the maps $\boldsymbol{a} : S \to C$ that they induce via their sections, that is,

$$\boldsymbol{a}(s) = \rho_a(s) \qquad \forall s \in S$$

Such maps, which can be seen as state-contingent consequences, are called (*Savage*) *acts*. Denote by $\boldsymbol{A}$ the collection of all of them. By (3), we can directly consider the preference $\succsim$ on $\boldsymbol{A}$ by setting $\boldsymbol{a} \succsim \boldsymbol{b}$ if and only if $a \succsim b$. The quartet $(\boldsymbol{A}, S, C, \succsim)$ represents the Savage decision problem, which can be viewed as a reduced form of the problem $(A, S, C, \rho, \succsim)$.

Although effective at a theoretical level, Savage acts may be somewhat artificial objects. As Marschak and Radner (1972, p. 13) remark, the notions in the quintet $(A, S, C, \rho, \succsim)$ "correspond more closely to the everyday connotations of the words" than do their Savage reduction $(\boldsymbol{A}, S, C, \succsim)$. For this reason, in applications it may be more natural to consider actions rather than acts (see, e.g., Section 4.9).

Among acts, bets play a special role because they elicit subjective probabilities, an insight due to de Finetti and Ramsey.[31] In particular, given any two consequences $c \succ c'$, we denote by cEc' the bet on event $E \subseteq S$ that pays the best consequence c if E obtains and pays c' otherwise. Given any two events E and F, a preference $cFc' \succsim cEc'$ reveals that the DM considers F (weakly) more likely than E.

2.5. Classical Decision Problems

As discussed in the Introduction, we suppose that the DM knows—because of his ex-ante structural information—that states are generated by a probability model $m \in \Delta(S)$

and so we take as a primitive the DMs' ranking of them, which subsumes their ranking of consequences. In any case, Consequentialism ensures that our modeling choice is consistent with such an alternative view.

30. As Savage (1954, p. 14) remarked, "If two different acts had the same consequences in every state of the world, there would from the present point of view be no point in considering them two different acts at all. An act is therefore identified with its possible consequences. Or, more formally, an act is a function attaching a consequence to each state of the world."

31. See Ramsey (1931) and de Finetti (1931, 1937). We refer to Gilboa and Marinacci (2013) for a discussion.

belonging to a given (finite) subset M of $\Delta(S)$.[32] Each m describes a possible generative mechanism. As such, it represents physical uncertainty, that is, the inherent randomness that states feature. In other words, the DM posits a model space M in addition to the state space S.[33] In so doing, he satisfies a central tenet of classical statistics à la Neyman–Pearson–Wald.[34] The model space might well be based on experts' advice, and its nonsingleton nature may reflect different advice.

Following Cerreia-Vioglio et al. (2013b), we take the "physical" information M as a primitive and enrich the standard Savage framework with this datum: the DM knows that the true model m generating the states belongs to the posited collection M. In terms of the basic preference $\succsim$, this translates into the requirement that betting behavior be consistent with datum M:

$$m(F) \geq m(E) \quad \forall m \in M \Longrightarrow cFc' \succsim cEc', \tag{4}$$

where cFc' and cEc' are bets on events F and E, with $c \succ c'$. If all models in M deem event F more likely than E, then the DM accordingly prefers betting on F to betting on E, that is, he deems F more likely than E.

The quintet $(A, S, C, M, \succsim)$ forms a Savage classical decision problem. In particular: for gamblers, models can be the (possible) composition of urns; for monetary authorities, the exogenous factors affecting the Phillips curve and the shocks' distributions; for firms, the prices' distributions; for investors, the returns' distributions; for civic officials, the peak ground accelerations' distributions; for engineers, the stiffness' distributions; for environmental policy makers, the distributions of climate sensitivity; and so on.

Urns. In a betting decision problem, suppose DMs know that the urn contains 90 balls, which can be either black or green or red. The state space is $S = \{B, G, R\}$ and so, without any further information, $M = \Delta(\{B, G, R\})$. If DMs are told that 30 balls are red, then $M = \{m \in \Delta(\{B, G, R\}) : m(R) = 1/3\}$. If instead they are told that half the balls are red, then $M = \{m \in \Delta(\{B, G, R\}) : m(R) = 1/2\}$; in this case, condition (4) implies that DMs are, for example, indifferent between bets on R and on $B \cup G$. Finally, if DMs are told the exact composition—say, with an equal number of each color—then $M = \{m\}$ is the singleton such that $m(B) = m(G) = m(R) = 1/3$.

Football. Consider a DM who has to bet on whether the local team will win a football match (see de Finetti 1977). If ties are not allowed, the state space is $S = \{W, L\}$. Suppose that, because of his ex-ante information (e.g., the two teams played this match

32. We call *probability models* the single probability measures m in the collection M. In statistics they are often called hypotheses, while the collection M itself is called a statistical model. Note that by assuming that the true model belongs to M, we abstract from misspecification issues.

33. For instance, a specification of M consists of all probabilities that belong to a suitable neighborhood of a posited benchmark model (Hansen and Sargent 2008, 2014).

34. See, for example, Wald (1950) and Neyman (1957); for a discussion, see Arrow (1951, p. 418). Stoye (2011) is a recent contribution to statistical decision theory from an axiomatic standpoint.

many times), the DM is able to assign the following probabilities that the local team will win under the different terrain conditions determined by the weather:

	Rainy	Cloudy	Sunny
Prob of W	$1/5$	$1/2$	$7/10$
Prob of L	$4/5$	$1/2$	$3/10$

Three models result, namely, $M = \{m_{\text{rainy}}, m_{\text{cloudy}}, m_{\text{sunny}}\} \subseteq \Delta(\{W, L\})$.

Health Insurance. As in Gilboa and Marinacci (2013), consider two DMs, John and Lisa, who have to decide whether to buy insurance against the risk of a heart disease. They are each 70 years old, are smokers, have no blood pressure problems, have a total cholesterol level of 310 mg/dL, with 45 mg/dL HDL-C (good cholesterol), and have systolic blood pressure of 130. What is the probability of a heart attack in the next ten years? John and Lisa consult a few web experts: using a "heart disease risk calculator" available on the web sites of several major hospitals, they obtain the following results:

Experts	John	Lisa
Mayo Clinic	25%	11%
National Cholesterol Education Program	27%	21%
American Heart Association	25%	11%
Medical College of Wisconsin	53%	27%
University of Maryland Heart Center	50%	27%

Thus the different experts, based on their data and medical models, provide quite different probability models for the event "heart attack in the next ten years" for each of the two DMs. In this case, John and Lisa end up with a set M consisting of four elements and three elements, respectively. Formally, if we set $s_1 =$ "heart attack in the next ten years" and $s_2 =$"no heart attack in the next ten years", the state space is $S = \{s_1, s_2\}$ and every probability model $m \in \Delta(S)$ is parameterized by the probability $m(s_1)$ that it assigns to s_1. Hence, we can write $M_{\text{John}} = \{25/100, 27/100, 53/100, 50/100\}$ and $M_{\text{Lisa}} = \{11/100, 21/100, 27/100\}$.

Environmental Issues. In environmental policy problems, probability distributions of climate sensitivity vary across different climate models proposed by different experts (see, e.g., Meinshausen et al. 2009; Rogelj, Meinshausen and Knutti 2012). The set M then consists of the collection of such distributions, as discussed by Millner, Dietz, and Heal (2013) and Heal and Millner (2014).

Population Games. Suppose that a static game is played recurrently in a stable environment by agents who, at each round, are drawn at random from large populations, with one population for each player role (Weibull 1996). In the game form $G = (I, (C_i)_{i \in I}, (A_i)_{i \in I}, (g_i)_{i \in I})$ the symbol I now denotes the set of player roles and $i \in I$ is the role of the agent drawn from population i. In that role the agent selects an action $a_i \in A_i$ which yields consequence $c_i = g(a_i, a_{-i})$ provided the other agents (the opponents) select actions a_{-i} in their roles.

For agents in role i the probability $\alpha_{-i} \in \Delta(A_{-i})$ describes a possible distribution of opponents' actions, with $\alpha_{-i}(a_{-i})$ being the fraction of opponents who select profile a_{-i} when drawn. The set $M_i \subseteq \Delta(A_{-i})$ is the collection of all these distributions that agents in role i posit.

Mixed Strategies. In an interactive situation among players who can commit to play actions selected by random devices, the probability $\alpha_{-i} \in \Delta(A_{-i})$ can be interpreted as a mixed strategy of the opponents of player i. If so, $\alpha_{-i}(a_{-i})$ becomes the probability that opponents' random devices select profile a_{-i} (in general, such devices are assumed to be independent across players). Here $M_i \subseteq \Delta(A_{-i})$ can be interpreted as the set of opponents' mixed strategies that player i considers.

3. Classical Subjective Expected Utility

3.1. Representation

Consider a Savage decision problem $(\boldsymbol{A}, S, C, M, \succsim)$. Cerreia-Vioglio et al. (2013b) show that a preference $\succsim$ satisfying Savage's axioms[35] and the consistency condition (4) is represented by the criterion $V : \boldsymbol{A} \to \mathbb{R}$ given by

$$V(\boldsymbol{a}) = \int_M \left(\int_S u(\boldsymbol{a}(s)) dm(s) \right) d\mu(m). \tag{5}$$

That is, the acts $\boldsymbol{a}$ and $\boldsymbol{b}$ are ranked as follows:

$$\boldsymbol{a} \succsim \boldsymbol{b} \iff V(\boldsymbol{a}) \geq V(\boldsymbol{b}).$$

Here $u : C \to \mathbb{R}$ is a von Neumann–Morgenstern utility function[36] that captures risk attitudes (i.e., attitudes toward physical uncertainty) and $\mu : 2^{M} \to [0, 1]$ is a subjective prior probability that quantifies the epistemic uncertainty about models, with support included in M. The subjective prior μ reflects some personal information on models that DMs may have, in addition to the structural information that allowed them to posit the collection M.[37] In particular, when that collection is based on the advice of

35. See, for example, Gilboa (2009, pp. 97–105).

36. That is, $c \succsim c'$ if and only if $u(c) \geq u(c')$.

37. Here we do not adopt the evidentialist view, known in economics as the Harsanyi doctrine (see, e.g., Aumann 1987; Morris 1995), that subjects with the same information should have the same subjective probabilities (for discussions, see Jaynes 2003, Chap. 12; Kelly 2008).

different experts, the prior may reflect the different weight (reliability) that DMs attach to each of them.

The quintet $(\boldsymbol{A}, S, C, M, \succsim)$ can therefore be represented in the form $(\boldsymbol{A}, S, C, M, u)$. Representation (5) may be called *classical subjective expected utility* because of the classical statistics tenet on which it relies. If we set

$$U(\boldsymbol{a}, m) = \int_S u(\boldsymbol{a}(s))dm(s)$$

we can then write the criterion as

$$V(\boldsymbol{a}) = \int_M U(\boldsymbol{a}, m)d\mu(m),$$

In words, the criterion considers the expected utility $U(\boldsymbol{a}, m)$ of each possible generative mechanism m and then averages them according to the prior μ. In some applications it is useful to write $V(\boldsymbol{a}, \mu)$ to emphasize the role of beliefs in an action's value.

This two-stage criterion can be seen as a decision-theoretic form of hierarchical Bayesian modeling.[38] As emphasized in the Introduction, all its probabilistic components are, ultimately, epistemic (and hence subjective) because they depend on some unmodeled background information known to DMs. If we denote such information by I, we can informally account for this key feature by writing criterion (5) in the following heuristic form:

$$V(\boldsymbol{a} \mid I) = \int_M \left(\int_S u(\boldsymbol{a}(s))dm(s|I) \right) d\mu(m \mid I).$$

Although we omit them for brevity, similar "information augmented" versions hold for the other criteria studied in this paper.[39]

Optimal acts solve the optimization problem $\max_{\boldsymbol{a} \in A} V(\boldsymbol{a})$. They thus depend on the preference $\succsim$ via the utility function u and the prior μ. To facilitate the presentation, hereafter we assume that optimal acts, if they exist, are unique (because, say, of suitable strict concavity assumptions on the objective function). We denote by $\hat{\boldsymbol{a}}$ the optimal act; to ease notation, we do not mark its dependence on u and μ (or on the set $\boldsymbol{A}$ of available acts).

Each prior μ induces a *predictive probability* $\bar{\mu} \in \Delta(S)$ through reduction:

$$\bar{\mu}(E) = \int_M m(E)d\mu(m) \qquad \forall E \in \Sigma. \tag{6}$$

38. See, for example, Bernardo and Smith (1994) and Berger (1993). For, m and s can be seen as realizations of two random variables—say, $\boldsymbol{m}$ and $\boldsymbol{s}$—with $m(s)$ a realization of the distribution of $\boldsymbol{s}$ conditional on $\boldsymbol{m}$ (see Picci 1977).

39. In these heuristic versions, conditioning on I is purely suggestive, without any formal meaning (to give useful content to I can be, indeed, a quite elusive problem) and without adopting, as just remarked, any evidentialist view.

In turn, the predictive probability allows us to rewrite the representation (5) as

$$V(\boldsymbol{a}) = U(\boldsymbol{a}, \bar{\mu}) = \int_S u(\boldsymbol{a}(s))d\bar{\mu}(s). \tag{7}$$

This reduced form of V is the original Savage subjective expected utility (SEU) representation.[40]

The predictive $\bar{\mu}$ is Savage's subjective probability, elicitable à la de Finetti–Ramsey via betting behavior on events. For, let $c, c' \in C$ be any two consequences, with $c \succ c'$. Without loss of generality, we can normalize u so that $u(c') = 0$ and $u(c) = 1$. If so, $cFc' \succsim cEc'$ if and only if $\bar{\mu}(F) \geq \bar{\mu}(E)$. Building on this simple remark allows one to show that the probability $\bar{\mu}$ may, in principle, be elicited (see, e.g., Gilboa 2009).

We remark that in the reduction operation that generates predictive probabilities, some important probabilistic features may disappear. For instance, in a binary state space $S = \{s_1, s_2\}$ consider the two collections $M = \{(0,1), (1,0)\}$ and $M' = \{(1/2 - \delta, 1/2 + \delta) : 0 \leq \delta \leq \varepsilon\}$, where $0 < \varepsilon < 1/2$ is an arbitrarily small quantity. Taking a uniform prior on each collection yields, in each case, the uniform predictive probability on S that assigns probability $1/2$ to each state. However, the collection M consists of two very different models whereas the collection M' consists of many almost identical models. Thus very different probabilistic scenarios are reduced to the same predictive probability.

Some special cases are important.

(i) If the support of μ is a singleton $\{m\}$ (i.e., $\mu = \delta_m$), then DMs believe, perhaps wrongly, that m is the true model. In this case the predictive probability trivially coincides with m and so criterion (5) reduces to the Savage expected utility criterion $U(\boldsymbol{a}, m)$. As a predictive probability, m is here a subjective probability (albeit a dogmatic one).

(ii) If M is a singleton $\{m\}$, then DMs have a maximal structural information and, consequently, know that m is the true model. There is no epistemic uncertainty, but only physical uncertainty (quantified by m).[41] Criterion (5) again reduces to the expected utility representation $U(\boldsymbol{a}, m)$, but now interpreted as a von Neumann–Morgenstern criterion since, absent epistemic uncertainty, subjective probabilities have no role to play. When combined with (7), this shows that classical SEU encompasses both the Savage and the von Neumann–Morgenstern representations.[42]

40. A further reduction, equation (20), will be discussed in Section 4.1. Note that probability measures in $\Delta(S)$ can play two different roles: predictive probabilities and probability models.

41. Knowledge of the true model ("known" probabilities) is a basic tenet of the rational expectations literature. Lucas (1977, p. 15) writes that "Muth (1961) . . . [identifies] . . . agents subjective probabilities . . . with 'true' probabilities, calling the assumed coincidence of subjective and 'true' probabilities *rational expectations*" (emphasis in the original).

42. Though the expected utility criterion was first proposed by Bernoulli (1738), the von Neumann and Morgenstern (1947) representation theorem marks the beginning of modern decision theory owing to its

(iii) If $M \subseteq \{\delta_s : s \in S\}$, then there is no physical uncertainty but only epistemic uncertainty (quantified by μ). We can identify prior and predictive probabilities: with a slight abuse of notation, we write $\mu \in \Delta(S)$ so that criterion (5) takes the form

$$V(\boldsymbol{a}) = \int_S u(\boldsymbol{a}(s))d\mu(s). \tag{8}$$

This is the form of the criterion that is relevant for decision problems without physical uncertainty.

(iv) If an act $\boldsymbol{a}$ is such that $U(\boldsymbol{a}, m) = U(\boldsymbol{a}, m')$ for all $m, m' \in \operatorname{supp} \mu$, we say that $\boldsymbol{a}$ is *crisp* (Ghirardato, Maccheroni, and Marinacci 2004). It is intuitive that crisp acts are not sensitive to epistemic uncertainty and that they feature the same physical uncertainty with respect to all models; hence, crisp acts can be regarded as purely risky acts.

Urns. In the previous urn example, suppose DMs are told that 30 balls are red, and so $M = \{m \in \Delta(\{B, G, R\}) : m(R) = 1/3\}$. This is the three-color problem of Ellsberg (1961). In order to parameterize M with the possible number $\theta \in \Theta = \{0, \ldots, 60\}$ of green balls, we denote by m_θ the element of M such that $m_\theta(G) = \theta/90$. Let $\boldsymbol{A} = \{\boldsymbol{a}_B, \boldsymbol{a}_G, \boldsymbol{a}_R\}$ be the 1 euro bets on the different colors. Suppose the prior μ is uniform—say, because DMs possess symmetric information about all possible compositions. If we normalize the utility function u by setting $u(1) = 1$ and $u(0) = 0$, then the following equalities hold:

$$V(\boldsymbol{a}_R) = \sum_{\theta=0}^{60} m_\theta(R)\mu(\theta) = \bar{\mu}(R) = \frac{1}{3}$$

$$V(\boldsymbol{a}_G) = \sum_{\theta=0}^{60} m_\theta(G)\mu(\theta) = \bar{\mu}(G) = \frac{1}{61}\sum_{\theta=0}^{60} \frac{\theta}{90} = \frac{1}{3}$$

$$V(\boldsymbol{a}_B) = \sum_{\theta=0}^{60} m_\theta(B)\mu(\theta) = \bar{\mu}(B) = 1 - \bar{\mu}(R \cup G) = \frac{1}{3}.$$

Therefore, $\boldsymbol{a}_R \sim \boldsymbol{a}_B \sim \boldsymbol{a}_G$. DMs are indifferent among the bets.

Now suppose that the DMs are instead given full information about the composition of the urn—say, that the colors are in equal proportion. Then $M = \{m\}$ is the singleton such that $m(G) = m(R) = m(B) = 1/3$ and we are back to the von Neumann–Morgenstern expected utility. Since $V(\boldsymbol{a}_R) = V(\boldsymbol{a}_G) = V(\boldsymbol{a}_B) = 1/3$, DMs are again indifferent among the bets. Despite the great difference in the quality of information, classical SEU leads to a similar preference pattern. In Section 4.4 we cast some doubts on the plausibility of this conclusion.

---

reliance on behavioral and so, in principle, testable axioms. But, there were some important earlier results on means and utilities (see Muliere and Parmigiani 1993).

Football. In the previous football example, let $\boldsymbol{a}$ and $\boldsymbol{b}$ be the 1 euro bets on the victory of, respectively, the local and guest teams. If we normalize the utility function u by setting $u(1) = 1$ and $u(0) = 0$, then

$$U(\boldsymbol{a}, m_{\text{rainy}}) = m_{\text{rainy}}(W)u(\boldsymbol{a}(W)) + m_{\text{rainy}}(L)u(\boldsymbol{a}(L)) = \frac{1}{5}u(1) + \frac{4}{5}u(0) = \frac{1}{5}$$
$$U(\boldsymbol{a}, m_{\text{cloudy}}) = m_{\text{cloudy}}(W)u(\boldsymbol{a}(W)) + m_{\text{cloudy}}(L)u(\boldsymbol{a}(L)) = \frac{1}{2}u(1) + \frac{1}{2}u(0) = \frac{1}{2}$$
$$\begin{aligned} U(\boldsymbol{a}, m_{\text{sunny}}) &= m_{\text{sunny}}(W)u(\boldsymbol{a}(W)) + m_{\text{sunny}}(L)u(\boldsymbol{a}(L)) \\ &= \frac{7}{10}u(1) + \frac{3}{10}u(0) = \frac{7}{10}. \end{aligned}$$

Consequently,

$$\begin{aligned} V(\boldsymbol{a}) &= \frac{1}{5}\mu(\text{rainy}) + \frac{1}{2}\mu(\text{cloudy}) + \frac{7}{10}\mu(\text{sunny}) \\ V(\boldsymbol{b}) &= \frac{4}{5}\mu(\text{rainy}) + \frac{1}{2}\mu(\text{cloudy}) + \frac{3}{10}\mu(\text{sunny}). \end{aligned}$$

Hence $\boldsymbol{a} \succsim \boldsymbol{b}$ if and only if $\mu(\text{rainy}) \leq (2/3)\mu(\text{sunny})$.

Monetary Policy. In the monetary policy decision problem, within a state $s = (w, \varepsilon, \theta)$ the pair (w, ε) represents random shocks and θ parameterizes a model economy. As in Battigalli et al. (2015a), we factor the probability models $m \in M \subseteq \Delta(W \times E \times \Theta)$ as follows:

$$m(w, \varepsilon, \theta) = q(w, \varepsilon) \times \delta_{\bar{\theta}}(\theta), \tag{9}$$

where $q \in \Delta(W \times E)$ is a shock distribution and $\delta_{\bar{\theta}} \in \Delta(\Theta)$ is a (Dirac) probability measure concentrated on a given economic model $\bar{\theta} \in \Theta$. Each model m thus corresponds to a shock distribution q and to a model economy θ.

Suppose that the distribution q of shocks is known, a common assumption in the rational expectations literature, so that there is epistemic uncertainty only about the structural component θ. In view of the factorization (9), we can then regard M as a subset of Θ and therefore define the prior μ directly on θ.[43] As a result, criterion (5) here becomes

$$\begin{aligned} V(\boldsymbol{a}) &= \int_M \left(\int_{W \times E \times \Theta} u(\boldsymbol{a}(w, \varepsilon, \theta)) dq(w, \varepsilon) \times \delta_{\bar{\theta}}(\theta) \right) d\mu(\bar{\theta}) \\ &= \int_M \left(\int_{W \times E} u(\boldsymbol{a}(w, \varepsilon, \theta)) dq(w, \varepsilon) \right) d\mu(\theta). \end{aligned}$$

43. Otherwise, μ should be defined on pairs (q, θ); that is, $\mu \in \Delta(\Delta(W \times E) \times \Theta)$.

The inner and outer integrals take care of, respectively, shocks' physical uncertainty and model economies' epistemic uncertainty.

Natural Hazards. In the natural hazard evacuation decision problem, suppose that officials posit M based on the advice of experts—say, seismologists' assessments of the PGA distribution (see, e.g., Baker 2008). If experts have different but dogmatic views, then $M \subseteq \{\delta_s : s \in S\}$: experts disagree but each of them has no doubts about the PGA caused by the upcoming earthquake. In this case there is no physical uncertainty, but only epistemic uncertainty.

Population Games. As we have remarked, in some applications it is more natural to consider actions rather than acts—that is, to consider the original decision problem $(A, S, C, \rho, \succsim)$. If so, criterion (5) takes the form $V(a) = \int_M (\int_S u(\rho(a,s))dm(s))d\mu(m)$. For instance, if in a population game we denote by μ_i the prior on $M_i \subseteq \Delta(A_{-i})$ held by agents in player role i, then we can write

$$V(a_i) = \int_{M_i} \left(\int_{A_{-i}} u(g_i(a_i, a_{-i}))d\alpha_{-i}(a_{-i}) \right) d\mu(\alpha_{-i})$$

for each action $a_i \in A_i$ (Battigalli et al. 2015a). Here the special case (iii) discussed previously becomes $M_i \subseteq \{\delta_{a_{-i}} : a_{-i} \in A_{-i}\}$; that is, the distributions of opponents' actions are degenerate. Agents drawn from a population then play the same action, so the population behaves as a single player. We are thus back to an interactive situation among standard players, in which the mass action interpretation is actually not relevant (e.g., an oligopoly problem). In this case we can write $V(a_i) = \int_{A_{-i}} u(g_i(a_i, a_{-i}))d\mu(a_{-i})$.

Financial Investment. The financial investment (or portfolio) decision problem is

$$\max_{a \in \Delta_{n-1}} \int_M \left(\int_S u(a \cdot s)dm(s) \right) d\mu(m)$$

if, to ease notation, we assume $w = 1$.[44] Its predictive form is

$$\max_{a \in \Delta_{n-1}} \int_S u(a \cdot s)d\bar{\mu}(s).$$

Suppose that there are two assets, a risk-free one with certain return r_f and a risky one with uncertain return r. In this case, the state space is the set R of all possible returns of the risky asset. If we denote by $a \in [0, 1]$ the fraction of wealth invested in the risky asset, then the problem becomes $\max_{a \in [0,1]} \int_M (\int_R u((1-a)r_f + ar)dm(r))d\mu(m)$.

44. Versions of this problem have been studied by, for example, Bawa, Brown, and Klein (1979), Kandel and Stambaugh (1996), Barberis (2000), and Pastor (2000).

Suppose $r - r_f = \beta x + (1-\beta)\varepsilon$, with $\beta \in [0,1]$. The scalar x can be interpreted as a predictor for the excess return, while ε is a random shock with distribution q. The higher is β, the more predictable is the excess return. Now, the state is $s = (\varepsilon, \beta)$, where ε and β are, respectively, its random and structural components. We assume, as in the previous monetary example, that $m(\varepsilon, \beta) = q(\varepsilon) \times \delta_{\bar{\beta}}(\beta)$, that is, $M \subseteq \Delta(\Delta(E) \times [0,1])$. Each model thus corresponds to a shock distribution and to a predictability structure. If the shock distribution q is known, the only unknown element is the predictability coefficient β: the investor is only uncertain about the predictability of the risky asset. The investment problem becomes

$$\max_{a \in [0,1]} \int_{[0,1]} \left(\int_E u(r_f + a(\beta x + (1-\beta)\varepsilon)) dq(\varepsilon) \right) d\mu(\beta). \tag{10}$$

In contrast, if the shock distribution and the predictability coefficient are both unknown, the problem is

$$\max_{a \in [0,1]} \int_{\Delta(E) \times [0,1]} \left(\int_E u(r_f + a(\beta x + (1-\beta)\varepsilon)) dq(\varepsilon) \right) d\mu(q, \beta). \tag{11}$$

In the terminology of Barberis (2000), problem (10) features only predictability uncertainty, while in problem (11) we have both parametric and predictability uncertainty.

A final remark. Models are often parameterized via a set Θ and a one-to-one map $\theta \longmapsto m_\theta$, so that we can write $M = \{m_\theta : \theta \in \Theta\}$ and $V(\boldsymbol{a}) = \int_\Theta (\int_S u(\boldsymbol{a}(s)) dm_\theta(s)) d\mu(\theta)$. The need for analytical tractability often leads researchers to use parameterizations that depend on specifying only a few coefficients (e.g., two for normal distributions). Yet some collections M of models admit a natural parameterization in which parameters have a concrete meaning, possibly in terms of (at least in principle) observables. So in the urn example we parameterized M with the possible number of green balls (i.e., $\Theta = \{0, \ldots, 60\}$), and in the football example with the weather conditions (i.e., $\Theta = \{\text{cloudy, rainy, sunny}\}$). In these cases, model uncertainty can be seen as uncertainty about such parameters, which in turn can be regarded as the (mutually exclusive and jointly exhaustive) contingencies that determine the variability of states (see de Finetti 1977; Pearl 1988, pp. 357–372). Different models in M thus account for different ways in which contingencies may affect this variability.[45]

3.2. Uniqueness

The support of the prior μ consists of probability models that are, in general, unobservable. For this reason, μ can be elicited only through hypothetical betting behavior (an "assisted form of introspection" according to Le Cam 1977). However,

45. The link between parameters and states can be formalized via conditioning if we introduce random variables that have s and θ as realizations (see footnote 38).

when in equation (5) the prior μ is unique, it can be elicited from the predictive $\bar{\mu}$ that it induces, which in turn is elicitable through bets on events. These considerations motivate our study in this section of μ uniqueness.

Linear Independence. The linear independence of M—not just its affine independence—underlies the desired uniqueness property. In particular, assume for simplicity that M and S are finite sets, with $M = \{m_1, \ldots, m_n\}$; linear independence means that, given any collection of scalars $\{\alpha_i\}_{i=1}^n$,

$$\sum_{i=1}^{n} \alpha_i m_i(s) = 0 \quad \forall s \in S \Longrightarrow \alpha_1 = \cdots = \alpha_n = 0.$$

This is a condition of linear independence of the n vectors $(m(s) : s \in S) \in \mathbb{R}^{|S|}$.

Cerreia-Vioglio et al. (2013b) elaborate on Teicher (1963) to show that μ is unique if M is linear independent. Hence, the reduction map $\mu \mapsto \bar{\mu}$ which through equation (6) relates predictive probabilities on the sample space to prior probabilities on space of models, is invertible on $\Delta(M)$. That is, distinct predictive probabilities $\bar{\mu}$ and $\bar{\mu}'$ correspond to distinct prior probabilities μ and μ' and since the Savagean predictive probability $\bar{\mu}$ can be elicited from betting behavior on events, it follows that any outside observer who is aware of M would then be able to infer the prior μ.

Orthogonality. Orthogonality is a simple but important sufficient condition for linear independence. Recall from Section 2.1 that, for a finite collection $M = \{m_1, \ldots, m_n\}$, this condition amounts to requiring the existence of a measurable partition $\{E_i\}_{i=1}^n$ of events such that, for each i, $m_i(E_i) = 1$ and $m_i(E_j) = 0$ if $j \neq i$. In words: for each model m_i, there is an element of the partition E_i that has probability 1 under that model and probability 0 under every other model.

In an intertemporal setup, this condition is satisfied by some fundamental classes of time series. Specifically, consider an intertemporal decision problem in which environment states are generated by a sequence of random variables $\{\tilde{z}_t\}$ defined on some (possibly unverifiable, except by Laplace's demon) underlying space and taking values on spaces Z_t that, for ease of exposition, we assume to be finite. For example, the sequence $\{\tilde{z}_t\}$ can model subsequent draws of balls from a sequence of (possibly identical) urns; here Z_t would consist of the possible colors of the balls that can be drawn from urn t.

Suppose, for convenience, that all spaces Z_t are finite and identical—each denoted by Z and endowed with the σ-algebra $\mathcal{B} = 2^Z$—and that the relevant state space S for the decision problem is the overall space $Z^\infty = \times_{t=1}^\infty Z_t = \times_{t=1}^\infty Z$. Its points $s = (z_1, \ldots, z_t, \ldots)$ are the possible paths generated by the sequence $\{\tilde{z}_t\}$. Without loss of generality, we identify $\{\tilde{z}_t\}$ with the coordinate process such that $\tilde{z}_t(s) = z_t$.

Endow Z^∞ with the product σ-algebra $\mathcal{B}^\infty$ generated by the elementary cylinders sets defined by $z^t = \{s \in Z^\infty : s_1 = z_1, \ldots, s_t = z_t\}$. The elementary cylinder sets are the basic events in this intertemporal setting. In particular, the sequence $\{\mathcal{B}^t\}$, called filtration—where $\mathcal{B}_0 \equiv \{S, \emptyset\}$ and $\mathcal{B}^t$ is the algebra generated by the cylinders z^t — records the building up of environment states. Clearly, $\mathcal{B}^\infty$ is the σ-algebra generated by the filtration $\{\mathcal{B}^t\}$.

In this intertemporal setup, then, the pair (S, Σ) is given by $(Z^\infty, \mathcal{B}^\infty)$. The set M of generative models consists of probability measures $m : \mathcal{B}^\infty \to [0, 1]$. Acts are adapted outcome processes $\boldsymbol{a} = \{\boldsymbol{a}_t\} : Z^\infty \to C$, often called *plans*. The consequence space C also has a product structure $C = \mathcal{C}^\infty$, where $\mathcal{C}$ is a common instant outcome space. More specifically, $\boldsymbol{a}_t(s) \in \mathcal{C}$ is the consequence at time t if state s obtains.

Criterion (5) here takes the form

$$V(\boldsymbol{a}) = \int_M \left(\int_{Z^\infty} u(\boldsymbol{a}(s))\, dm(s) \right) d\mu(m). \tag{12}$$

Under standard conditions, the intertemporal utility function $u : \mathcal{C}^\infty \to \mathbb{R}$ has a classic discounted form $u(c_1, \ldots, c_t, \ldots) = \sum_{\tau=1}^{\infty} \beta^{\tau-1} \upsilon(c_\tau)$, with subjective discount factor $\beta \in [0, 1]$ and (bounded) instantaneous utility function $\upsilon : \mathcal{C} \to \mathbb{R}$.

As is well known (see, e.g., Billingsley 1965, p. 39), models are orthogonal in the stationary and ergodic case, which includes the standard independent and identically distributed (i.i.d.) setup as a special case. Formally, we have the following proposition.

PROPOSITION 1. *A finite collection M of models that make the process $\{\tilde{z}_t\}$ stationary and ergodic is orthogonal.*

So in this fundamental case, (12) features a cardinally unique utility function u and a unique prior μ, with $\operatorname{supp}\mu \subseteq M$. Since a version of Proposition 1 holds also for collections of homogeneous Markov chains, we can conclude that time series models widely used in applications satisfy the orthogonality conditions that ensure the uniqueness of prior μ. Classical SEU thus provides a framework for empirical work that relies on such time series (as is often the case in the finance and macroeconomics literatures).

Finally, note that the Savage reduced form $V(\boldsymbol{a}) = \int_{Z^\infty} u(\boldsymbol{a}(s))\, d\bar{\mu}(s)$ of the representation (12) features a predictive probability $\bar{\mu}$ that makes the process $\{\tilde{z}_t\}$ stationary when, as in the previous result, models are stationary and ergodic (exchangeable, in the special i.i.d. case). The ergodic theorem (and the de Finetti representation theorem) can be seen as providing conditions when a converse holds.

Asymptotic Contingencies. In the asymptotic setting of this section, parameters may be given a concrete interpretation, in the spirit of the discussion at the end of the previous section. For, given a collection of models $M = \{m_\theta\}_{\theta \in \Theta}$ in parametric form, under some probabilistic invariance conditions on the process $\{\tilde{z}_t\}$ (see, e.g., Diaconis and Freedman 1987) there exists a sequence of $\mathcal{B}^t$-measurable functions $\varphi_t : Z^\infty \to \Theta$ such that $m_\theta(E_\theta) = 1$ for every θ, where $E_\theta = \{(z_1, \ldots, z_t, \ldots) \in Z^\infty : \lim_{t \to +\infty} \varphi_t(z^t) = \theta\}$. For instance, if models make exchangeable a process $\{\tilde{z}_t\}$ of 0s and 1s, then $m_\theta(z^t) = \prod_{\tau=1}^{t} \theta^{z_\tau} (1 - \theta)^{(1 - z_\tau)}$ and $\varphi_n(z^t) = \sum_{\tau=1}^{t} z_\tau / t$. The events E_θ are the contingencies that determine the models m_θ in terms of the asymptotic behavior of the functions φ_t. Model uncertainty is thus uncertainty about the events E_θ—that is, about the asymptotic behavior of the functions φ_t.[46] This interpretation of parameters in terms of "asymptotic" contingencies is sometimes called *predictive* because it originates with the de Finetti representation theorem and the predictive approach it pioneered (see Cifarelli and Regazzini 1996).

3.3. Stochastic Consequences

In some applications, actions are assumed to deliver consequences that are stochastic and not (as we have assumed so far) deterministic.[47] Lotteries on consequences, rather than "pure" consequences, thus become the outcomes of actions when states obtain.

To see how to accommodate stochastic consequences in our setting, we must temporarily abandon the Savage setting and consider an action based classical decision problem $(A, S, C, \rho, M, \succsim)$. Assume that the state space has the Cartesian structure $S = S_1 \times S_2$ and that all posited models $m \in M$ can be factored as

$$m(s_1, s_2) = \tilde{m}(s_1)\gamma(s_2), \tag{13}$$

where $\gamma \in \Delta(S_2)$ is a distribution on S_2 that is known to the DM and so is common across all posited models. The stochastic consequence function $\tilde{\rho} : A \times S_1 \to \Delta(C)$ defined by

$$\tilde{\rho}(a, s_1)(c) = \gamma(s_2 : \rho(a, s_1, s_2) = c) \tag{14}$$

associates a lottery $\tilde{\rho}(a, s_1) \in \Delta(C)$ on consequences with each pair (a, s_1). By a simple change of variable, for each action a we have

$$\begin{aligned} V(a) &= \int_M \left(\int_S u(\rho(a,s)) dm(s) \right) d\mu(m) \\ &= \int_M \left(\int_{S_1} \left(\int_{S_2} u(\rho(a, s_1, s_2)) d\gamma(s_2) \right) d\tilde{m}(s_1) \right) d\mu(\tilde{m}) \\ &= \int_M \left(\int_{S_1} \left(\int_C u(c)\tilde{\rho}(a, s_1)(c) \right) d\tilde{m}(s_1) \right) d\mu(\tilde{m}), \end{aligned}$$

where $\int_C u(c)\tilde{\rho}(a, s_1)(c)$ is the expected utility of lottery $\tilde{\rho}(a, s_1) \in \Delta(C)$. Hence, under the factorization (13) we can equivalently rank actions through either the deterministic consequence function ρ on the full state space S or the stochastic consequence function $\tilde{\rho}$ on the partially specified state space S_1, where models differ.

46. Since $m_\theta(E_\theta) = 1$ for every θ, the existence of the functions $\{\varphi_t\}$ requires that models $\{m_\theta\}_{\theta \in \Theta}$ be orthogonal (the relations between these functions and orthogonal models are studied by Mauldin et al. 1983). In turn (paraphrasing Ornstein and Weiss 1990) this ensures that with probability 1 a single sampling of process $\{\tilde{z}_t\}$ suffices to determine the process exactly (of course, such a sampling is an idealized notion that involves an actual infinity of observations).

47. This section was written in collaboration with Pierpaolo Battigalli.

The previous argument provides a simple foundation in our setting for classical decision problems $(A, \tilde{S}, \Delta(C), \tilde{\rho}, \tilde{M}, \succsim)$ featuring stochastic consequence functions $\tilde{\rho}$ and partially specified state spaces $\tilde{S}$ that do not exhaust all payoff-relevant uncertainty.[48] Yet because the specification of an exhaustive state space is often difficult, it may be convenient in some applications to work directly with stochastic consequence functions, viewing them as a purely distributional modeling of the residual uncertainty not accounted for in the state space at hand. If so, a converse of the previous derivation shows that there is an underlying augmented state space, with the same Cartesian structure, on which an equivalent deterministic consequence function can be defined. Specifically, set $S_1 = \tilde{S}$, $S_2 = C^{A\times S}$ (i.e., S_2 is the collection of all maps $s_2 : A \times S \to C$), and $\rho(a, s_1, s_2) = s_2(a, s_1)$. Finally, if we define $\gamma(s_2) = \times_{(a,s)\in A\times S}\tilde{\rho}(a, s)(s_2(a, s))$, then M is the collection of all m on $S_1 \times S_2$ defined via equation (13). Condition (14) is easily seen to hold and so actions can be equivalently ranked through either $\tilde{\rho}$ or ρ.[49] However, this augmented state space is just a formal construction. Stochastic consequence functions are a more parsimonious and direct modeling tool.

Under a stochastic version of Consequentialism, stochastic consequence functions $\tilde{\rho}$ induce maps $\boldsymbol{a} : S \to \Delta(C)$ from states to lotteries defined by $\boldsymbol{a}(s) = \tilde{\rho}_a(s)$, which are called *Anscombe–Aumann acts*. Given any consequence set C, the set $\Delta(C)$ is always convex. So, unlike Savage acts, Anscombe–Aumann acts have a natural convex structure—an analytical feature most useful in establishing axiomatic foundations of decision criteria (see Gilboa and Marinacci 2013).

That said, in what follows we continue to study deterministic consequence functions. Our analysis does still apply also to stochastic consequence function, but we omit details for brevity.

4. Within the Bayesian Paradigm: Smooth Ambiguity

4.1. Representation

Monetary Case. Suppose that acts are monetary, that is, the consequence space C is an interval of the real line.[50] Representation (5) can then be equivalently written as[51]

$$V(\boldsymbol{a}) = \int_M (u \circ u^{-1})(U(\boldsymbol{a}, m))d\mu(m) = \int_M u(c(\boldsymbol{a}, m))d\mu(m), \tag{15}$$

48. See Fishburn (1981). Machina (2004) introduces a state space in the real line that, by building on some insights of Poincaré, makes it possible to define almost-objective events. Mutatis mutandis, along with the Lebesgue measure they provide a counterpart of the common γ measure.

49. We can regard $\tilde{\rho}$, $s_2 \in C^{A\times S}$, and γ as, respectively, the behavioral, pure, and mixed strategies of a chance player that plays after observing (a, s). Then, the construction of the augmented state space is an application of the realization equivalence theorem of Kuhn (1953).

50. In what follows, whenever we consider monetary acts we tacitly assume that C is an interval of the real line and that $u : C \to \mathbb{R}$ is strictly increasing and continuous.

51. Note that $\operatorname{Im} u$ is an interval in the real line because u is strictly increasing and continuous.

where $c(\boldsymbol{a}, m)$ is the certainty equivalent

$$c(\boldsymbol{a}, m) = u^{-1}(U(\boldsymbol{a}, m)) \tag{16}$$

of act $\boldsymbol{a}$ under model m. The profile $\{c(\boldsymbol{a}, m) : m \in \operatorname{supp} \mu\}$ represents the scope of model uncertainty, that is, the epistemic uncertainty that the DM experiences when dealing with alternative possible probabilistic models m that may generate states. In particular,

$$V(\boldsymbol{a}) = \int_M u(c(\boldsymbol{a}, m)) d\mu(m)$$

is the decision criterion that the DM uses to address that epistemic uncertainty, while

$$U(\boldsymbol{a}, m) = \int_S u(\boldsymbol{a}(s)) dm(s)$$

is how the DM addresses the physical uncertainty that each model m features.

Implicitly, the representation (5) thus assumes identical attitudes toward physical and epistemic uncertainty, both modeled by the same function u. But, there is no reason to expect that this is generally the case; for instance, DMs might well be more averse to epistemic than to physical uncertainty. In the next section, the celebrated Ellsberg paradox will starkly illustrate this issue.

It is therefore important to generalize the classical SEU representation (5) by distinguishing such attitudes. Toward that end, we adapt the *smooth ambiguity* model of Klibanoff, Marinacci, and Mukerji (2005, hereafter KMM) to the present setup.[52] Specifically, acts are ranked according to the criterion

$$V(\boldsymbol{a}) = \int_M (v \circ u^{-1})(U(\boldsymbol{a}, m)) d\mu(m) \tag{17}$$

$$= \int_M v(c(\boldsymbol{a}, m)) d\mu(m). \tag{18}$$

Here $c(\boldsymbol{a}, m)$ is the certainty equivalent (16) of act $\boldsymbol{a}$ under model m, and both u and μ are interpreted as in the classical SEU representation. Moreover, in light of (18), the strictly increasing and continuous function $v : \boldsymbol{C} \to \mathbb{R}$ represents attitudes toward model uncertainty—that is, toward the epistemic uncertainty that DMs experience when dealing with alternative possible probabilistic models m that may generate states. A negative attitude toward model uncertainty is modeled by a concave v, which can be interpreted as aversion to (mean preserving) spreads in the certainty equivalents $c(\boldsymbol{a}, m)$. Note that u and v share the same monetary domain $\boldsymbol{C}$.

52. Related, yet distinct, models have been proposed by Segal (1987), Davis and Paté-Cornell (1994), Nau (2006), Ergin and Gul (2009), and Seo (2009).

General Case. Set $\phi = v \circ u^{-1} : \operatorname{Im} u \subseteq \mathbb{R} \to \mathbb{R}$,[53] so that we can rewrite (17) as

$$V(\boldsymbol{a}) = \int_M \phi(U(\boldsymbol{a}, m))d\mu(m) \tag{19}$$

or, in shorthand notation,

$$V(\boldsymbol{a}) = \mathrm{E}_\mu \phi(\mathrm{E}_m u(\boldsymbol{a})).$$

This is the original formulation of KMM, who show that a notion of ambiguity aversion is captured by the concavity of ϕ (KMM, p. 1862). Because the domain $\operatorname{Im} u$ of ϕ is denominated in utils, such concavity can be interpreted as aversion to (mean preserving) spreads in expected utilities (KMM, p. 1851). By standard results in risk theory, the concavity of ϕ amounts to v being more concave than u—that is, to the existence of a strictly increasing and concave function g such that $v = g \circ u$. Hence ambiguity aversion in the present two-stage setup amounts to a higher degree of aversion toward epistemic than toward physical uncertainty.

Ambiguity neutrality corresponds to a linear ϕ—that is, to equal attitudes toward physical and epistemic uncertainty. If so, $u = v$ and we return to the classical SEU representation (5).

We close with an important observation: representation (19) continues to hold even when C is not an interval of the real line, but just any set. Hence, equation (19) can be seen as the extension of the certainty equivalent representation (17) to general settings.

Portability. Representation (19) is "portable" across decision problems because it parameterizes personality traits: risk attitudes given by the function u and ambiguity attitudes given by the function ϕ. Such traits can be assumed to be constant across decision problems (with monetary consequences); in contrast, state spaces and beliefs change according to the situation. Therefore, once u and ϕ are elicited or calibrated (say, in the laboratory), the representation can be used for different problems. As in traditional expected utility analysis, u can be elicited through lotteries; ϕ can be elicited through uncertain prospects based on some canonical mechanism (such as urns) that simulates model uncertainty.

Sources Matter. At a methodological level, the smooth ambiguity model drops a basic feature of Savage's original approach, a feature that is often described by saying that his approach was able "to reduce uncertainty to risk". To see why this is the case, assume for simplicity that the consequence space C is finite. Then the SEU criterion (7) can be equivalently written as

$$V(\boldsymbol{a}) = \sum_{c \in C} u(c)\bar{\mu}_{\boldsymbol{a}}(c), \tag{20}$$

53. See KMM (p. 1859). Since u is strictly increasing, the inverse u^{-1} exists; since u is continuous, the image $\operatorname{Im} u = \{u(c) : c \in C\}$ is an interval of the real line. Note that ϕ inherits from u and v the properties of being continuous and strictly increasing.

where $\bar{\mu}_{\boldsymbol{a}}$, the probability that act $\boldsymbol{a}$ induces on consequences, is defined by

$$\bar{\mu}_{\boldsymbol{a}}(c) = \bar{\mu}(s \in S : s \in \boldsymbol{a}^{-1}(c)) \qquad \forall c \in C.$$

The expected utility representation (20) is a "risk reduction" of the SEU criterion (7), that is often used in applications. The probability $\bar{\mu}_{\boldsymbol{a}}$ is a lottery in the jargon of risk theory, and (20) shows that SEU relies, in the final analysis, on the lotteries induced by acts. These lotteries are evaluated per se, à la von Neumann–Morgenstern, independently of acts and of their underlying state spaces and subjective probabilities. In fact, consider acts $\boldsymbol{a} : S \to C$ and $\boldsymbol{a}' : S' \to C$ defined on different state spaces: as long as the subjective probabilities $\bar{\mu}$ and $\bar{\mu}'$ on such spaces induce the same lottery (i.e., $\bar{\mu}_{\boldsymbol{a}} = \bar{\mu}'_{\boldsymbol{a}'}$), the SEU criterion (7) treats the two acts identically despite the altogether different kind of uncertainty that they might feature. This invariance across decision problems featuring different sources of uncertainty is a remarkable property of SEU.

The smooth representation partly abandons this property. We distinguished two sources of uncertainty, physical and epistemic. The former is featured by each model $m \in M$, each of which induces a lottery on consequences defined by $m_{\boldsymbol{a}}(c) = m(s \in S : s \in \boldsymbol{a}^{-1}(c))$. The latter is characterized by a prior μ on M that induces a lottery on consequences defined by $\mu_{\boldsymbol{a}}(c) = \mu(m \in M : c(\boldsymbol{a}, m) = c)$. Risk reductionism would require these lotteries to be treated per se, à la von Neumann–Morgenstern, independently of their underlying sources of uncertainty. Accordingly, if u is the relevant von Neumann–Morgenstern utility, then we should use it for all these lotteries and consider the expected utilities $\sum_{c \in C} u(c) m_{\boldsymbol{a}}(c)$ and $\sum_{c \in C} u(c) \mu_{\boldsymbol{a}}(c)$. This is, indeed, what the SEU criterion (15) does. In contrast, the smooth ambiguity criterion (17) permits different attitudes toward the two uncertainty sources via their respective von Neumann–Morgenstern utility functions u and v and expected utilities $\sum_{c \in C} u(c) m_{\boldsymbol{a}}(c)$ and $\sum_{c \in C} v(c) \mu_{\boldsymbol{a}}(c)$. In this sense, (17) is a source dependent criterion. Within each source, however, it keeps the invariance principle of SEU (inter alia, this helps make the criterion analytically tractable and portable, as remarked previously).

Thus, in this setup we maintain a Bayesian approach by using a single probability measure to quantify probabilistic judgments within each source of uncertainty. However, different confidence in such judgments (whatever feature of a source causes it) translate as different degrees of aversion to uncertainty across sources, and so in different von Neumann–Morgenstern utility functions.[54] The next example illustrates this key point.

A Coin Illustration. Betting on coins is, intuitively, greatly affected by whether or not the coins are well tested. Heads and tails are judged to be equally likely when betting on a well-tested coin that has been flipped a number of times with approximately equal

54. See Chew and Sagi (2008), Abdellaoui et al. (2011), and Gul and Pesendorfer (2015) for recent decision models that explicitly consider sources of uncertainty. Smith (1969) is an early paper that drops source independence and considers different uncertainty attitudes for different sources of uncertainty.

instances of heads and tails. There is only physical uncertainty, no model uncertainty. When dealing with an untested coin, however, we have both physical and model uncertainty; here the different models correspond to different possible biases of the coin. Suppose that DMs consider (if only for symmetry) a uniform prior over such models; then once again heads and tails are judged to be equally likely by the resulting predictive probability, as will be seen momentarily. Yet the evidence behind such judgments—and thus the confidence in them—is dramatically different. Hence, DMs may well prefer, ceteris paribus, betting on tested coins rather than betting on untested ones.[55]

This preference naturally emerges in our setting by taking into account the negative attitude of DMs toward model uncertainty. Specifically, call I the tested coin and II the untested one. Actions a_{I} and a_{II} are, respectively, bets of 1 euro on coin I and on coin II. The state space is

$$S = \{H, T\} \times \{H, T\} = \{HH, HT, TH, TT\}.$$

Here state HH obtains when both coins land heads, state HT obtains when coin I lands heads and coin II lands tails, and so forth. The consequence function is

$$\rho(a_{\mathrm{I}}, s) = \begin{cases} 1 & \text{if } s \in HH \cup HT \\ 0 & \text{if } s \in TH \cup TT \end{cases} \quad ; \quad \rho(a_{\mathrm{II}}, s) = \begin{cases} 1 & \text{if } s \in HH \cup TH \\ 0 & \text{if } s \in HT \cup TT. \end{cases}$$

The following table summarizes the decision problem in terms of acts:

	HH	HT	TH	TT
$\boldsymbol{a}_{\mathrm{I}}$	1	1	0	0
$\boldsymbol{a}_{\mathrm{II}}$	1	0	1	0

Given the available information, it is natural to set

$$M = \left\{ m \in \Delta(s) : m(HH \cup HT) = m(TH \cup TT) = \frac{1}{2} \right\}.$$

In words, M consists of all models that assign probability $1/2$ to either outcome for the tested coin I. No specific probability is, instead, assigned to the outcome of the untested coin. As a result, act $\boldsymbol{a}_{\mathrm{I}}$ is unaffected by model uncertainty, while $\boldsymbol{a}_{\mathrm{II}}$ is heavily affected.

If we normalize the utility function by setting $u(1) = 1$ and $u(0) = 0$, then

$$V(\boldsymbol{a}_{\mathrm{I}}) = \int_M \phi(m(HH \cup HT)) d\mu(m) = \phi\left(\frac{1}{2}\right)$$

55. In the jargon of epistemology, the evidence has the same balance but different weight (see Kelly 2008; it is well known that the "weight of evidence" notion traces back to Keynes 1921). Gilboa (2009) discusses the importance of this coin example in the genesis of David Schmeidler's approach to ambiguity (see Section 5). An insightful discussion can be found in Jaynes (1959, pp. 184–185).

and

$$V(\boldsymbol{a}_{\mathrm{II}}) = \int_M \phi(m(HH \cup TH))d\mu(m).$$

Because the prior μ is unaffected by model uncertainty, it turns out to be irrelevant for the value of act $\boldsymbol{a}_{\mathrm{I}}$. It matters, instead, for act a_{II}. Suppose μ is uniform, so that all models are equally weighted. If so, the value of act $\boldsymbol{a}_{\mathrm{II}}$ is given by the Riemann integral $V(\boldsymbol{a}_{\mathrm{II}}) = \int_0^1 \phi(x)dx$. If ϕ is strictly concave, then it follows from the Jensen inequality that

$$V(\boldsymbol{a}_{\mathrm{II}}) = \int_0^1 \phi(x)dx < \phi\left(\int_0^1 xdx\right) = \phi\left(\frac{1}{2}\right) = V(\boldsymbol{a}_{\mathrm{I}}).$$

This ranking captures the desired preference for betting on the tested coin. It emerges through the explicit modeling of model uncertainty (and its aversion) that the smooth criterion permits.

We close by observing that for the predictive probability we have $\bar{\mu}(HH \cup HT) = \bar{\mu}(TH \cup TT) = 1/2$ as well as

$$\bar{\mu}(HH \cup TH) = \int_M m(HH \cup TH)d\mu(m) = \int_0^1 xdx = 1/2$$
$$\bar{\mu}(HH \cup TH) = \int_M m(HH \cup TH)d\mu(m) = \int_0^1 (1-x)dx = 1/2.$$

The predictive probabilities are thus not able to distinguish the different weights that characterize the evidence about the two coins. This result confirms our remark in Section 3.1: in the reduction operation that generates predictive probabilities, some important information may be lost.

A Final Twist. For decision problems in which physical uncertainty is not relevant, $M \subseteq \{\delta_s : s \in S\}$ and so supp $\mu \subseteq \{\delta_s : s \in S\}$. Representation (17) then takes the form $V(\boldsymbol{a}) = \int_S v(\boldsymbol{a}(s))d\mu(s)$. When $v = u$ (identical attitudes toward physical and epistemic uncertainty), we are back in the situation of equation (8). Otherwise, when $v \neq u$ (attitudes differ), it is the function v that matters.[56]

4.2. Analytical Examples

The two-stage representations that we have seen are very general. Some standard specifications of its components would be helpful if we seek to compare findings

56. If we consider stochastic consequences (Section 3.3) with Anscombe–Aumann acts $\boldsymbol{a} : S \to \Delta(C)$, the special case $\int_S \phi(u(\boldsymbol{a}(s)))d\mu(s)$ of representation (19) when supp $\mu \subseteq \{\delta_s : s \in S\}$ is sometimes called second-order (or recursive) expected utility (see Neilson 2010).

across different applications. Hence we next consider a few specifications of utility functions.[57]

(i) Suppose that both utility functions u and v are CARA,[58] with $u(x) = -e^{-\alpha x}$ and $v(x) = -e^{-\beta x}$ for all $x \in \mathbb{R}$. Then $\phi : (-\infty, 0) \to \mathbb{R}$ is given by $\phi(y) = (v \circ u^{-1})(y) = -(-y)^{\beta/\alpha}$ for each $y < 0$, and so, in shorthand notation,

$$V(\boldsymbol{a}) = -\mathrm{E}_{\mu} \left(\mathrm{E}_{m} e^{-\alpha \boldsymbol{a}}\right)^{\beta/\alpha}.$$

In particular, ϕ is concave provided that $\alpha \leq \beta$. If $\alpha = \beta$, we return to the SEU criterion $V(\boldsymbol{a}) = -\mathrm{E}_{\bar{\mu}} e^{-\alpha \boldsymbol{a}}$. If $u(x) = x$ (i.e., if u is risk neutral), then $\phi = v$ and so

$$V(\boldsymbol{a}) = -\mathrm{E}_{\mu} e^{-\beta \mathrm{E}_{m} \boldsymbol{a}}. \tag{21}$$

(ii) Suppose that both utility functions u and v are CRRA,[59] with $u(x) = x^{\alpha}$ and $v(x) = x^{\beta}$ for all $x > 0$. Then $\phi : (0, +\infty) \to \mathbb{R}$ is given by $\phi(y) = (v \circ u^{-1})(y) = y^{\beta/\alpha}$ for each $y > 0$, so

$$V(\boldsymbol{a}) = \mathrm{E}_{\mu} \left(\mathrm{E}_{m} \boldsymbol{a}^{\alpha}\right)^{\beta/\alpha}.$$

In particular, ϕ is concave provided that $\alpha \geq \beta$ (in which case ϕ as well is CRRA). If $\alpha = \beta$, we return to the SEU criterion $V(\boldsymbol{a}) = \mathrm{E}_{\bar{\mu}} \boldsymbol{a}^{\alpha}$. If $v(x) = \log x$, then $\phi = \alpha^{-1} v$ and so $V(\boldsymbol{a}) = \alpha^{-1} \mathrm{E}_{\mu} \log(\mathrm{E}_{m} \boldsymbol{a}^{\alpha})$.

(iii) A CARA ϕ, say $\phi(x) = -e^{-\lambda x}$, exhibits constant absolute ambiguity aversion (see KMM p. 1866). In this case,

$$V(\boldsymbol{a}) = -\mathrm{E}_{\mu} e^{-\lambda \mathrm{E}_{m} u(\boldsymbol{a})}. \tag{22}$$

Since $v = \phi \circ u$, we have $v(x) = -e^{-\lambda u(x)}$; that is, the function v is "CARA in utils" as opposed to the previous "CARA in money" case $v(x) = -e^{-\beta x}$. In the risk-neutral case $u(x) = x$, the two notions coincide; in particular, (21) and (22) are equal.

As KMM (p. 1886) argues, a constant Arrow–Pratt index $-\phi''/\phi'$ captures a form of constant (absolute) ambiguity aversion (CAAA), characterized by an exponential ϕ. In a similar vein, the notions of DAAA and IAAA can be introduced. Finally, a power ϕ can be viewed as capturing a form of CRAA.[60]

---

57. As to probabilities, normality is often assumed on models (though there is an increasing emphasis on non-normal models); priors can be standardized through objective Bayesian methods, which often rely on noninformative priors (see, e.g., Kass and Wasserman 1996; Berger 2006).

58. A utility function u is CARA (constant absolute risk aversion) if, for all $x \in \mathbb{R}$, either $u(x) = x$ or $u(x) = -e^{-\alpha x}$ for some $\alpha > 0$.

59. A utility function u is CRRA (constant relative risk aversion) if, for all $x > 0$, either $u(x) = \log x$ or $u(x) = x^{\alpha}$ for some $0 < \alpha < 1$.

60. Grant and Polak (2013) discuss constant ambiguity attitudes in terms of the weak certainty independence axiom of Maccheroni, Marinacci, and Rustichini (2006). Their analysis has been extended to DAAA by Xue (2012), who suitably weakens that independence axiom and axiomatizes a constant superadditive DAAA version of variational preferences. Relative attitudes are instead related to the

4.3. Certainty Equivalents

In the smooth ambiguity model, monetary acts $\boldsymbol{a}$ have two certainty equivalents:

$$c(\boldsymbol{a},\mu) = v^{-1}\left(\int_M (v \circ u^{-1})(U(\boldsymbol{a},m))d\mu(m)\right)$$

and

$$\mathsf{c}(\boldsymbol{a},\mu) = \phi^{-1}\left(\int_M \phi(U(\boldsymbol{a},m))d\mu(m)\right).$$

The former expression is measured in money but the latter uses utils—the only kind of certainty equivalent that is well defined for general (i.e., not only monetary) acts. Since $\mathsf{c}(\boldsymbol{a},\mu) = u(c(\boldsymbol{a},\mu))$, it follows that the two expressions are related by a simple change of scale. In any case, it is the monetary certainty equivalent $c(\boldsymbol{a},\mu)$ that corresponds to the standard notion of certainty equivalent and that generalizes the certainty equivalent $c(\boldsymbol{a},m)$. It features some noteworthy properties.

PROPOSITION 2. *For each monetary act* $\boldsymbol{a}$ *and each prior* μ*, the following statements hold:* [61]

(i) $c(k,\mu) = k$ *for each* $k \in \mathbb{R}$*;*
(ii) $c(\boldsymbol{a}+k,\mu) = c(\boldsymbol{a},\mu) + k$ *for each* $k \in \mathbb{R}$*, provided* u *and* v *are CARA;*
(iii) $c(\alpha\boldsymbol{a},\mu) = \alpha c(\boldsymbol{a},\mu)$ *for each* $\alpha \in [0,1]$*, provided* u *and* v *are CRRA;*
(iv) $c(\cdot,\mu)$ *is concave provided* ϕ *is concave and* u *and* v *are either both CARA or both CRRA.*

Since $c(\boldsymbol{a},\bar{\mu}) = u^{-1}(U(\boldsymbol{a},\bar{\mu}))$, the Jensen inequality implies that for each act $\boldsymbol{a}$ and each prior μ the inequality $c(\boldsymbol{a},\bar{\mu}) \geq c(\boldsymbol{a},\mu)$ holds provided ϕ is concave. Ambiguity aversion thus lowers the value of acts (in Section 4.9 we discuss whether this results in a more cautious choice behavior).

We close the section by computing the certainty equivalents for the examples given in the previous section.

EXAMPLE 1. (i) In the CARA example of Section 4.2(i), we have $v^{-1}(x) = -(1/\beta)\log(-x)$ and $\phi^{-1}(x) = -(-x)^{\alpha/\beta}$. Hence, in shorthand notation,

$$c(\boldsymbol{a},\mu) = -\frac{1}{\beta}\log \mathrm{E}_\mu\left(\mathrm{E}_m e^{-\alpha\boldsymbol{a}}\right)^{\beta/\alpha} \quad \text{and} \quad \mathsf{c}(\boldsymbol{a},\mu) = -\left(\mathrm{E}_\mu\left(\mathrm{E}_m e^{-\alpha\boldsymbol{a}}\right)^{\beta/\alpha}\right)^{\alpha/\beta}.$$

multiplicative version of the weak certainty independence axiom studied by Chateauneuf and Faro (2009).

61. With a slight abuse of notation, k denotes both a scalar and a constant act (e.g., $(\boldsymbol{a}+k)(s) = \boldsymbol{a}(s) + k$ for each $s \in S$).

Under risk neutrality,

$$\mathsf{c}(\boldsymbol{a},\mu) = c(\boldsymbol{a},\mu) = -\frac{1}{\beta}\log \mathrm{E}_{\mu} e^{-\beta \mathrm{E}_m \boldsymbol{a}}. \tag{23}$$

(ii) In the CRRA example of Section 4.2(ii), we have $v^{-1}(x) = x^{1/\beta}$ and $\phi^{-1}(x) = x^{\alpha/\beta}$. Hence

$$c(\boldsymbol{a},\mu) = \left(\mathrm{E}_{\mu}\left(\mathrm{E}_m \boldsymbol{a}^{\alpha}\right)^{\beta/\alpha}\right)^{1/\beta} \quad \text{and} \quad \mathsf{c}(\boldsymbol{a},\mu) = \left(\mathrm{E}_{\mu}\left(\mathrm{E}_m \boldsymbol{a}^{\alpha}\right)^{\beta/\alpha}\right)^{\alpha/\beta}.$$

(iii) In the CARA ϕ example of Section 4.2(iii), we have $v^{-1}(x) = u^{-1}(-(1/\lambda)\log(-x))$ and $\phi^{-1}(x) = -(1/\lambda)\log(-x)$. Therefore,

$$c(\boldsymbol{a},\mu) = u^{-1}\left(-\frac{1}{\lambda}\log \mathrm{E}_{\mu} e^{-\lambda \mathrm{E}_m u(\boldsymbol{a})}\right) \quad \text{and} \quad \mathsf{c}(\boldsymbol{a},\mu) = -\frac{1}{\lambda}\log \mathrm{E}_{\mu} e^{-\lambda \mathrm{E}_m u(\boldsymbol{a})}.$$

■

The last utility certainty equivalent $\mathsf{c}(\boldsymbol{a},\mu)$ can be seen as a version of the multiplier preferences of Hansen and Sargent (2001, 2008); under risk neutrality, it coincides with equation (23). We will return to this important case in Section 5.2.

4.4. Illustration: Ellsberg Paradoxes

The classic paradoxes presented by Ellsberg (1961) starkly illustrate the importance of distinguishing between attitudes toward physical and epistemic uncertainty. The first paradox involves two colors in two urns, the second paradox (which we have already encountered) involves three colors in one urn.

Two Urns. Consider two urns, I and II, with 100 balls in each.

1. The DM is told that:
   (i) in both urns, balls are either white or black;
   (ii) in urn I, there are 50 black and 50 white balls.
2. The DM is given no information on the proportion of white and black balls in urn II.

Thus urn I features only physical uncertainty whereas urn II features both physical and epistemic uncertainties. The DM must choose from the following 1 euro bets on the colors of a ball drawn from each urn:

1. bets $\boldsymbol{a}_{\mathrm{I}}$ and $\boldsymbol{b}_{\mathrm{I}}$, which pay 1 euro if the ball drawn from urn I is, respectively, black or white;
2. bets $\boldsymbol{a}_{\mathrm{II}}$ and $\boldsymbol{b}_{\mathrm{II}}$, which pay 1 euro if the ball drawn from urn II is, respectively, black or white.

We can easily model such choices in a Savage framework. The state space is

$$S = \{B, W\} \times \{B, W\} = \{BB, BW, WB, WW\},$$

where state BB obtains when a black ball is drawn from both urns, state BW obtains when a black ball is drawn from urn I and a white ball is drawn from urn II, and so forth. The decision problem is summarized by the following table.

	BB	BW	WB	WW
$\boldsymbol{a}_{\mathrm{I}}$	1	1	0	0
$\boldsymbol{b}_{\mathrm{I}}$	0	0	1	1
$\boldsymbol{a}_{\mathrm{II}}$	1	0	1	0
$\boldsymbol{b}_{\mathrm{II}}$	0	1	0	1

Given the available information, it is natural to set

$$M = \left\{ m \in \Delta(s) : m(BB \cup BW) = m(WW \cup WB) = \frac{1}{2} \right\}.$$

In words, M consists of all models that give probability $1/2$ to the drawing of either color from the "physical" urn I. In contrast, the models in M do not assign any specific probability to the drawing of either color from the "epistemic" urn II. As a result, M consists of 101 elements. We can parameterize them with the possible number $\theta \in \Theta = \{0, \ldots, 100\}$ of black balls in urn II denoting by m_θ the element of M such that $m_\theta(BB \cup WB) = 1 - m_\theta(BW \cup WW) = \theta/100$.

Suppose the DM ranks these bets according to the classical SEU criterion. First of all, normalize the utility function by setting $u(1) = 1$ and $u(0) = 0$. Then

$$U(\boldsymbol{a}_{\mathrm{I}}) = U(\boldsymbol{b}_{\mathrm{I}}) = \frac{1}{2}.$$

Now suppose, given the symmetry of information about urn II, that the prior μ is uniform. Then[62]

$$U(\boldsymbol{a}_{\mathrm{II}}) = \sum_{\theta \in \Theta} m_\theta(BB \cup WB)\mu(\theta) = \frac{1}{101} \sum_{\theta=0}^{100} \frac{\theta}{100} = \frac{1}{2}$$

and a similar argument shows that $U(\boldsymbol{b}_{\mathrm{II}}) = 1/2$. We conclude that, according to the classical SEU model, the DM should be indifferent among all bets; that is, $\boldsymbol{a}_{\mathrm{I}} \sim \boldsymbol{b}_{\mathrm{I}} \sim \boldsymbol{a}_{\mathrm{II}} \sim \boldsymbol{b}_{\mathrm{II}}$. It is, however, plausible that the DM would rather bet on the physical urn I than on the epistemic urn II:

$$\boldsymbol{a}_{\mathrm{I}} \sim \boldsymbol{b}_{\mathrm{I}} \succ \boldsymbol{a}_{\mathrm{II}} \sim \boldsymbol{b}_{\mathrm{II}}. \tag{24}$$

62. An early version of the calculations presented here can be found in Boole (1854, pp. 370–375).

Besides introspection, overwhelming experimental evidence confirms this preference pattern. Remarkably, the pattern emerges as soon as we distinguish—via the smooth ambiguity model—between attitudes toward physical and epistemic uncertainty. To this end, we normalize also the utility function v by setting $v(1) = 1$ and $v(0) = 0$, and we assume that $\phi = v \circ u^{-1}$ is strictly concave, so that the DM is strictly ambiguity averse (i.e., he is strictly more averse to epistemic than to physical uncertainty). We still have $V(\boldsymbol{a}_{\mathrm{I}}) = V(\boldsymbol{b}_{\mathrm{I}}) = 1/2$, but now

$$V(\boldsymbol{a}_{\mathrm{II}}) = \sum_{\theta \in \Theta} \phi(m_\theta(BB \cup WB))\mu(\theta) = \frac{1}{101} \sum_{\theta=0}^{100} \phi\left(\frac{\theta}{100}\right)$$
$$< \phi\left(\frac{1}{101} \sum_{\theta=0}^{100} \frac{\theta}{100}\right) = \phi\left(\frac{1}{2}\right).$$

A similar argument shows that $V(\boldsymbol{b}_{\mathrm{II}}) = V(\boldsymbol{a}_{\mathrm{II}})$. Hence the Ellsberg pattern (24) emerges as soon as we distinguish the different attitudes by considering a higher aversion to epistemic than to physical uncertainty.

Three Colors. We return now to Section 3.1's urn example in which DMs were told that 30 balls are red, so that $M = \{m \in \Delta(\{B, G, R\}) : m(R) = 1/3\}$. We observed that classical SEU implies that $\boldsymbol{a}_R \sim \boldsymbol{a}_B \sim \boldsymbol{a}_G$—that is, the DMs are indifferent among betting on any of the three colors—when a uniform prior μ is assumed. Such a prior results from the symmetric information about all possible compositions of green and blue balls.

Given the available information, it is plausible that DMs prefer betting on red to betting on either green or blue; that is,

$$\boldsymbol{a}_R \succ \boldsymbol{a}_B \sim \boldsymbol{a}_G. \tag{25}$$

In fact, all posited models agree on the probability of R, but they differ substantially on that of the other two colors.

We adopt the normalizations of the two urns paradox, that is, $u(1) = v(1) = 1$ and $u(0) = v(0) = 0$. Under the uniform prior of Section 3.1, we have $V(\boldsymbol{a}_R) = \phi(1/3)$ and

$$V(\boldsymbol{a}_G) = \sum_{\theta=0}^{60} \phi(m_\theta(G))\mu(\theta) = \frac{1}{61} \sum_{\theta=0}^{60} \phi\left(\frac{\theta}{90}\right) < \phi\left(\frac{1}{61} \sum_{\theta=0}^{60} \frac{\theta}{90}\right) = \phi\left(\frac{1}{3}\right).$$

A similar argument shows that $V(\boldsymbol{a}_G) = V(\boldsymbol{a}_B)$. Hence pattern (25) is easily justified once we account for differences in attitudes.

4.5. Illustration: Structural Reliability (and Mean-Preserving Spreads)

Consider a general structural reliability problem that generalizes the cantilever beam example of Section 2.2.[63] An engineer has to select the design of a structure within a set A of possible alternative designs. The reliability of each design depends on an exogenous state $s \in S$, which can be seen as the realization of an underlying random variable. A *limit state function* $g : A \times S \to \mathbb{R}$ identifies, for each design a, the failure event $F_a = \{s \in S : g(a,s) < 0\}$ and the safe event $F_a^c = \{s \in S : g(a,s) \geq 0\}$.[64] In the cantilever beam example, $g(a,s) = d - \tau(a,s)$. As in that example, the consequence function $\rho : A \times S \to \mathbb{R}$ is given by

$$\rho(a,s) = \begin{cases} \delta + c(a) & \text{if} \quad s \in F_a \\ c(a) & \text{if} \quad s \in F_a^c, \end{cases}$$

where δ is the damage cost of failure and $c(a)$ is the cost of design a.[65] The induced act $\boldsymbol{a}$ is thus binary:

$$\boldsymbol{a}(s) = \begin{cases} \delta + c(a) & \text{if} \quad s \in F_a \\ c(a) & \text{if} \quad s \in F_a^c. \end{cases}$$

Since $c(a) > c(a) + \delta$, act $\boldsymbol{a}$ can be viewed as a bet on the safe event F_a^c. Hence, the preference for design a over design b can be seen as a preference to bet on the safe event F_a^c rather than on the safe event F_b^c.

Given a model $m \in \Delta(S)$, the *failure probability* m_a of design a is $m_a = m(F_a)$. The engineer does not know the true model, but he is able to posit a set M of possible models.[66] He has prior probability μ over such models and ranks designs through the smooth ambiguity criterion

$$\begin{aligned} V(\boldsymbol{a}) &= \int_M \phi\left(\int_S u(\boldsymbol{a}(s))dm(s)\right) d\mu(m) \\ &= \int_M \phi((u(\delta + c(a)) - u(c(a)))m_a + u(c(a)))d\mu(m). \end{aligned}$$

If ϕ is linear, we obtain the classical SEU criterion (see equation (7))

$$U(\boldsymbol{a},\bar{\mu}_a) = (u(\delta + c(a)) - u(c(a)))\bar{\mu}_a + u(c(a)),$$

63. This section was written in collaboration with Veronica Cappelli. Borgonovo et al. (2015) show how some well-known models of risk analysis (e.g., Kaplan and Garrick 1981; Kaplan, Perla, and Bley 1983) that deal with technological uncertainty can be embedded in the setup of this paper.

64. The failure event is often described as $F_a = \{s \in S : g(a,s) \leq 0\}$. For our purposes, however, it does not matter whether the event is identified by a weak or a strict inequality. Moreover, for simplicity we assume that g is scalar and not vector valued (because of multiple components that the structure might have) and that there is a single threshold, zero, and not multiple ones (because of different possible damage levels).

65. To ease notation, we assume that costs δ and $c(a)$ are negative scalars.

66. For simplicity, we assume that the limit state function g is known.

where $\bar{\mu}_a$ is the *mean failure probability* of design a which is given by $\bar{\mu}_a = \int_M m(F_a) d\mu(m)$.

Under the classical SEU criterion, an engineer should be indifferent between two designs that share the same mean failure probability and the same costs. Yet, as Der Kiureghian (2008, p. 354) remarks:[67] "This result appears somewhat counter-intuitive, as with identical costs and mean failure probabilities, one would expect a preference for the case with smaller uncertainty in the failure probability estimate. But simple derivations [...] show that aversion to uncertainty in failure probability is not logical."

We show that such counterintuitive indifference is not so much a logical necessity as a consequence of the failure to distinguish attitudes toward physical and toward epistemic uncertainty. Once this distinction is made and a nonlinear ϕ is considered, the counterintuitive indifference no longer holds. To see why this is the case, consider a decision problem $(\boldsymbol{A}, S, C, \succsim)$ in which the preference $\succsim$ is represented by the smooth ambiguity criterion (19)—namely, $V(\boldsymbol{a}) = \int_M \phi(U(\boldsymbol{a}, m)) d\mu(m)$ where we assume (for simplicity) that $\operatorname{supp} \mu$ is a finite set.[68] In particular, by (7) the equality $V(\boldsymbol{a}) = U(\boldsymbol{a}, \bar{\mu})$ holds when ϕ is linear.

We say that act $\boldsymbol{b}$ is a (*subjective*) *mean-preserving spread*[69] of act $\boldsymbol{a}$ if $U(\boldsymbol{a}, \bar{\mu}) = U(\boldsymbol{b}, \bar{\mu})$ and, given that $|M| > 2$, there are models $m', m'' \in \operatorname{supp} \mu$ such that

$$U(\boldsymbol{b}, m'') \leq U(\boldsymbol{a}, m'') \leq U(\boldsymbol{a}, m') \leq U(\boldsymbol{b}, m') \tag{26}$$

and

$$U(\boldsymbol{a}, m) = U(\boldsymbol{b}, m) \qquad \forall m \neq m', m''. \tag{27}$$

By conditions (26) and (27), the two acts share the same expected utilities for all models except m' and m'', where act $\boldsymbol{b}$ features expected utilities that are more spread out than those of $\boldsymbol{a}$. As a result, $\boldsymbol{b}$ exhibits a variability in expected utility that is no less than the corresponding variability of $\boldsymbol{a}$.

The condition $U(\boldsymbol{a}, \bar{\mu}) = U(\boldsymbol{b}, \bar{\mu})$ disciplines such a comparison of variability by restricting it to pairs of acts that share the same expected utility under the predictive probability $\bar{\mu}$. This condition amounts to requiring acts $\boldsymbol{a}$ and $\boldsymbol{b}$ to be indifferent when ϕ is linear (i.e., when acts are ranked according to classical SEU). However, the greater variability of act $\boldsymbol{b}$ might well reduce its appeal relative to act $\boldsymbol{a}$. The next general result shows that this is indeed the case as long as ϕ is concave, that is, provided the DM is more averse toward epistemic than toward physical uncertainty or, equivalently, if he is ambiguity averse.

PROPOSITION 3. *If the DM is ambiguity averse, then $V(\boldsymbol{b}) \leq V(\boldsymbol{a})$ whenever act $\boldsymbol{b}$ is a mean preserving spread of act $\boldsymbol{a}$.*

67. Der Kiureghian assumes linear utility, but doing so is not essential for his argument.

68. Recall from Section 4.1 that $U(\boldsymbol{a}, m) = \int_S u(\boldsymbol{a}(s)) dm(s)$.

69. This notion adapts a definition of Ghirardato and Marinacci (2001), which was in turn based on the classic work of Rothschild and Stiglitz (1970).

In short, the indifference prescribed by classical SEU between acts $\boldsymbol{a}$ and $\boldsymbol{b}$ fails to hold as soon as the distinction between the attitudes toward epistemic and toward physical uncertainty is made.[70]

Returning now to the design choice, consider two designs a and b that share the same costs. Since $U(\boldsymbol{a}, m) = (u(\delta + c(a)) - u(c(a)))m_a + u(c(a))$, it easily follows that $\boldsymbol{b}$ is a mean-preserving spread of $\boldsymbol{a}$ if and only if $\bar{\mu}_a = \bar{\mu}_b$ and there are $m', m'' \in \operatorname{supp} \mu$ such that

$$m''_b \leq m''_a \leq m'_a \leq m'_b$$

and $m_a = m_b$ for all $m \neq m', m''$. By Proposition 3, ambiguity-averse engineers prefer design a over design b even though the two designs share the same mean failure probability and the same costs. In other words, engineers who are more averse toward epistemic than toward physical uncertainty exhibit the "preference for the case with smaller uncertainty in the failure probability estimate" that Der Kiureghian (2008), previously quoted, claims is intuitive.

4.6. Extreme Ambiguity Aversion: Maxmin

Wald. Under extreme ambiguity aversion—that is, when ambiguity aversion "goes to infinity"—the smooth ambiguity criterion (17) reduces in the limit (under finiteness of M) to the classic Wald (1950) maxmin criterion[71]

$$V(\boldsymbol{a}) = \min_{m \in M} \int_S u(\boldsymbol{a}(s)) dm(s). \tag{28}$$

Under this very cautious criterion, the DM "maxminimizes" over all possible probability models in M. Prior probabilities do not play any role.

We illustrate this limit result in an important special case. Mathematically, increasing ambiguity aversion corresponds to increasing concavity of the function $\phi : \operatorname{Im} u \to \mathbb{R}$. When this function is twice differentiable, its Arrow–Pratt index of concavity is $\lambda_\phi = -\phi''/\phi'$. In the CARA case $\phi_\lambda(x) = -e^{-\lambda x}$, for instance, the parameter $\lambda > 0$ is the constant Arrow–Pratt index, and so ϕ_λ features constant absolute ambiguity aversion (Section 4.2(iii)). Consider the utility certainty equivalent (Example 1)

$$\mathsf{c}_\lambda(\boldsymbol{a}, \mu) = -\frac{1}{\lambda} \log \int_M e^{-\lambda \int_S u(\boldsymbol{a}(s)) dm(s)} d\mu(m) \tag{29}$$

with $\lambda > 0$. Here ambiguity aversion goes to infinity or to zero provided that $\lambda \to +\infty$ or $\lambda \to 0$, respectively. In particular, since $v_\lambda = \phi_\lambda \circ u$, it is the aversion to epistemic uncertainty that underlies these limit cases.

70. We will refrain (for the sake of brevity) from further analysis, but all these considerations hold also for the case of acts that—though not directly comparable through a mean-preserving spread—can be connected via a finite sequence of intermediate acts that are, instead, comparable in that sense (see Rothschild and Stiglitz 1970, p. 231).

71. See KMM (p. 1867). This section was written in collaboration with Simone Cerreia-Vioglio.

It can be easily shown that, if $\operatorname{supp} \mu = M$, then

$$\lim_{\lambda \to 0} \mathtt{C}_\lambda(\boldsymbol{a}, \mu) = \int_S u(\boldsymbol{a}(s)) d\bar{\mu}(s) \quad \text{and}$$

$$\lim_{\lambda \to +\infty} \mathtt{C}_\lambda(\boldsymbol{a}, \mu) = \min_{m \in M} \int_S u(\boldsymbol{a}(s)) dm(s).$$

In other words, as aversion to ambiguity (and so to epistemic uncertainty) goes either to infinity or to zero, we return either to the classical SEU criterion or to the Wald maxmin criterion.

The latter case is especially interesting: the Wald criterion emerges as the result of extreme aversion to epistemic uncertainty. This finding sheds further light on the extremity of such criterion.

A Grain of Subjectivity. More generally, when it only holds the inclusion $\operatorname{supp} \mu \subseteq M$, we have

$$\lim_{\lambda \to +\infty} \mathtt{C}_\lambda(\boldsymbol{a}, \mu) = \min_{m \in \operatorname{supp} \mu} \int_S u(\boldsymbol{a}(s)) dm(s).$$

In this case, a grain of genuine subjectivity remains: the prior μ plays the role of selecting which models in M are relevant, and so must be included in the minimization. The prior simply classifies models as "in" and "out" and makes no further, finer, assessment on their likelihood. The limit criterion

$$V(\boldsymbol{a}) = \min_{m \in \operatorname{supp} \mu} \int_S u(\boldsymbol{a}(s)) dm(s) \tag{30}$$

can be interpreted as a Waldean version of the classic Gilboa–Schmeidler decision model (to be discussed in Section 5). In applications, this is often the version of the model that is used.

Statewise Maxmin. Suppose that the support of μ coincides with the collection of all probability measures on some event E; that is, $\operatorname{supp} \mu = \Delta(E)$. Then

$$V(\boldsymbol{a}) = \min_{m \in \operatorname{supp} \mu} \int_S u(\boldsymbol{a}(s)) dm(s) = \min_{s \in E} u(\boldsymbol{a}(s)). \tag{31}$$

In words: when the support consists of all models conceivable on some event, then the maxmin criterion takes a statewise form (on the event). DMs consider only the utilities that they can achieve state by state, probabilities (of any sort) play no role.[72]

72. This nonprobabilistic criterion may be relevant in what are known as decision problems under ignorance, in which information is too limited for any kind of probabilistic representation. Milnor (1954), Arrow and Hurwicz (1972), and Cohen and Jaffray (1980) study this important class of decision problems.

A No-Trade Illustration. In a frictionless financial market, consider a primary asset $y : S \to \mathbb{R}$ that pays out $y(s)$ if state $s \in S$ obtains. The market price of each unit of the asset is p. Investors have to decide how many units $x \in \mathbb{R}$ of the asset to trade. If $x > 0$, investors buy the asset; if $x < 0$, they sell it; if $x = 0$, there is no trade.

In any case, a portfolio with x units of the asset has a, state contingent, payoff $yx - px$.[73] The associated act $\boldsymbol{x} : S \to \mathbb{R}$ is given by $\boldsymbol{x}(s) = y(s)x - px$. There is no trade on asset y when $V(\boldsymbol{x}) < V(\mathbf{0})$ for all $x \neq 0$, that is, when investors do not benefit from either buying or from selling the asset. Assume risk neutrality. The next result, which adapts to the present setup a classic result of Dow and Werlang (1992), shows that with maxmin behavior there might be no trade (unless the asset is crisp[74]).

PROPOSITION 4. *Under maxmin behavior* (29), *there is no trade in asset y if and only if its price satisfies*

$$\min_{m \in \operatorname{supp} \mu} \mathrm{E}_m(y) < p < \max_{m \in \operatorname{supp} \mu} \mathrm{E}_m(y). \tag{32}$$

Condition (32) requires $\operatorname{supp} \mu$ to be nonsingleton; thus model uncertainty is essential for the result. In particular: the larger is the support of μ (and so the perceived model uncertainty), the larger the inequality and hence the set of prices that result in no trade.

Extreme ambiguity aversion may thus freeze market transactions—an important insight that has been often discussed in the literature.[75] This suggests that sufficiently high ambiguity aversion may arbitrarily reduce trade volumes. To see why this is the case, suppose investors rank these acts through the utility certainty equivalent (29). Under risk neutrality, we then have, in shorthand notation,

$$\mathsf{c}_\lambda(\boldsymbol{x}, \mu) = -\frac{1}{\lambda} \log \mathrm{E}_\mu e^{-\lambda \mathrm{E}_m \boldsymbol{x}}. \tag{33}$$

The next simple property will be useful.

LEMMA 1. $\mathsf{c}_\lambda(\boldsymbol{x}, \mu) \leq \mathsf{c}_{-\lambda}(\boldsymbol{x}, \mu)$ *for all* $x \in \mathbb{R}$, *with equality if* $\boldsymbol{x}$ *is crisp.*

We say that there is *at most ε-trade volume* on asset y if $\mathsf{c}_\lambda(\boldsymbol{x}, \mu) < 0$ for all $x \notin (-\varepsilon, \varepsilon)$; in words, investors may benefit only from transactions involving quantities smaller than ε. Next we show that sufficiently high ambiguity aversion can explain any arbitrarily low trade volume (unless the asset is crisp).

PROPOSITION 5. *Let* $\varepsilon > 0$. *There is at most ε-trade volume in asset y if*

$$\mathsf{c}_{\varepsilon\lambda}(y, \mu) < p < \mathsf{c}_{-\varepsilon\lambda}(y, \mu). \tag{34}$$

73. We abstract from any intertemporal issue (see Section 4.10).

74. That is, $\mathrm{E}_m \boldsymbol{x} = \mathrm{E}_{m'} \boldsymbol{x}$ for all $m, m' \in \operatorname{supp} \mu$.

75. See, for example, Guidolin and Rinaldi (2013).

In particular, the inequality holds when $\varepsilon\lambda$ is large enough (something that requires a higher λ if ε is smaller) provided the previous no-trade inequality (32) holds.[76] So whenever the price of an asset precludes its being traded under maxmin behavior, a higher ambiguity aversion corresponds to lower trade volume on the asset under criterion (33).

Extreme Risk Aversion. So far we have regarded maxmin behavior as the result of extreme ambiguity aversion—that is, of a much higher aversion to model uncertainty than to risk. Formally, however, the outcome statewise maxmin criterion à la Leontief $V(\boldsymbol{a}) = \min_{s\in S} \boldsymbol{a}(s)$ can be seen as a limit version of the classic SEU criterion (5) as risk aversion becomes increasingly larger (see Laffont 1989). In other words, the Leontief criterion can correspond either to extreme risk aversion or to risk neutrality with extreme model uncertainty aversion. The latter interpretation is consistent with presuming that DMs are more averse to model uncertainty than to risk. Nonetheless, variations of the previous no trade (resp., low trade) results hold under extreme (resp., high enough) levels of risk aversion. It remains an empirical and experimental issue to determine which is the more plausible behavioral hypothesis: high risk aversion or normal risk aversion and high model uncertainty aversion.

4.7. *Quadratic Approximation: Mean-Variance*

For monetary acts, the smooth ambiguity criterion admits a simple quadratic approximation that generalizes the well-known quadratic approximation of expected utility. To show this we take a consequence w and an act $\boldsymbol{b}$,[77] and we consider the certainty equivalent

$$c(w + \boldsymbol{b}, \mu) = v^{-1}(\mathrm{E}_\mu v(c(w + \boldsymbol{b}, m))) \equiv v^{-1}\left(\int_M v(c(w + \boldsymbol{b}, m))d\mu(m)\right).$$

Maccheroni, Marinacci, and Ruffino (2013) show that, under standard differentiability assumptions on the functions u and v, the quadratic approximation of this certainty equivalent is given by[78]

$$c(w + \boldsymbol{b}, \mu) \approx w + \mathrm{E}_{\bar{\mu}}\boldsymbol{b} - \frac{1}{2}\lambda_{\boldsymbol{u}}(w)\sigma^2_{\bar{\mu}}(\boldsymbol{b}) - \frac{1}{2}(\lambda_v(w) - \lambda_{\boldsymbol{u}}(w))\sigma^2_\mu(\mathrm{E}(\boldsymbol{b})), \quad (35)$$

76. Since $\lim_{\alpha\to+\infty} c_\alpha(y, \mu) = \min_{m\in \mathrm{supp}\,\mu} \mathrm{E}_m(y)$ and $\lim_{\alpha\to+\infty} c_{-\alpha}(y, \mu) = \max_{m\in \mathrm{supp}\,\mu} \mathrm{E}_m(y)$.

77. With a slight abuse of notation, we denote by w the consequence and also the constant act equal to w in all states. In the financial applications that originally motivated quadratic approximations, w is interpreted as initial wealth and $\tilde{b}$ as an investment.

78. In (35) and thereafter, $\lambda_u = -u''/u'$ and $\lambda_v = -v''/v'$ are the classic Arrow–Pratt coefficients of risk aversion.

where $\mathrm{E}(\boldsymbol{b}) : M \to \mathbb{R}$ is the random variable

$$m \mapsto \mathrm{E}_m \boldsymbol{b} \equiv \int_S \boldsymbol{b}(s) dm(s)$$

that associates the expected value of act $\boldsymbol{b}$ under each possible model m, and $\sigma^2_\mu(\mathrm{E}(\boldsymbol{b}))$ is its variance. The quadratic approximation (35) extends the classic Arrow–Pratt version

$$c(w + \boldsymbol{b}, \mu) \approx w + \mathrm{E}_{\bar{\mu}} \boldsymbol{b} - \frac{1}{2} \lambda_u(w) \sigma^2_{\bar{\mu}}(\boldsymbol{b}) \tag{36}$$

of the SEU criterion (7). Relative to the Arrow–Pratt approximation, (35) features the novel term

$$\frac{1}{2}(\lambda_v(w) - \lambda_u(w)) \sigma^2_\mu(\mathrm{E}(\boldsymbol{b})).$$

This term is an *ambiguity premium* jointly determined by the variance $\sigma^2_\mu(\mathrm{E}(\boldsymbol{b}))$, an information trait that captures the scope of model uncertainty that the DM perceives, and by the difference $\lambda_v(w) - \lambda_u(w)$, a taste trait that captures the DM's differences in attitudes toward physical and toward epistemic uncertainty.[79] In this regard, note that $\lambda_v \geq \lambda_u$ when v is more concave than u, that is, when the DM is more averse to epistemic than to physical uncertainty.

The approximation (35) leads to a natural generalization of the classic mean-variance model in which DMs rank acts $\boldsymbol{a}$ in terms of the robust mean-variance functional given by

$$\mathrm{E}_{\bar{\mu}} \boldsymbol{a} - \frac{\lambda}{2} \sigma^2_{\bar{\mu}}(\boldsymbol{a}) - \frac{\theta}{2} \sigma^2_\mu(\mathrm{E}(\boldsymbol{a})), \tag{37}$$

where λ and θ are positive coefficients. This criterion is determined by the three parameters λ, θ, and μ. In particular, when $\theta = 0$ we return to the usual mean-variance functional. In light of equation (35), criterion (37) can be viewed as a local quadratic approximation of a smooth ambiguity criterion at a constant w such that $\lambda = \lambda_u(w)$ and $\theta = \lambda_v(w) - \lambda_u(w)$. Thus, the taste parameters λ and θ model the DM's attitudes toward physical and epistemic uncertainty, respectively. Higher values of these parameters correspond to stronger negative attitudes.

The information parameter μ determines the variances $\sigma^2_{\bar{\mu}}(\boldsymbol{a})$ and $\sigma^2_\mu(\mathrm{E}(\boldsymbol{a}))$ that measure, respectively, the physical and epistemic uncertainties that the DM perceives when evaluating act $\boldsymbol{a}$. Higher values of these variances correspond to a DM's poorer information regarding such uncertainties.

Since the variance $\sigma^2_{\bar{\mu}}(\boldsymbol{b})$ can be decomposed by the law of total variance,[80] the approximation (35) can be rearranged according to the Arrow–Pratt coefficients of u and v:

$$c(w + \boldsymbol{b}, \mu) \approx w + \mathrm{E}_{\bar{\mu}} \boldsymbol{b} - \frac{\lambda_u(w)}{2} \mathrm{E}_\mu \sigma^2(\boldsymbol{b}) - \frac{\lambda_v(w)}{2} \sigma^2_\mu(\mathrm{E}(\boldsymbol{b})). \tag{38}$$

79. See Jewitt and Mukerji (2011) and Lang (2015) for related notions of ambiguity premium.

80. That is, $\sigma^2_{\bar{\mu}}(\boldsymbol{b}) = E_\mu(\sigma^2(\boldsymbol{b})) + \sigma^2_\mu(E(\boldsymbol{b}))$.

This formulation suggests the conditions under which such an approximation is exact. It is well known that, if u is CARA and $\boldsymbol{b}$ has a normal distribution with mean k and variance σ^2, then the Arrow–Pratt approximation (36) takes the following exact form:

$$c(w + \boldsymbol{b}, \mu) = w + k - \frac{1}{2}\lambda_u(w)\sigma^2.$$

This result easily generalizes to the quadratic approximation (38): if both u and v are CARA utility functions with respective constant Arrow–Pratt coefficients α and β, and if $\boldsymbol{b}$ has a normal distribution with unknown mean k and known variance σ^2, then (38) takes the exact form

$$c(w + \boldsymbol{b}, \mu) = w + \bar{k} - \frac{1}{2}\alpha\sigma^2 - \frac{1}{2}\beta\sigma_\mu^2 \tag{39}$$

as long as the prior on the unknown mean k is given by a normal distribution with parameters $\bar{k}$ and σ_μ^2.[81]

Finally, the quadratic approximation may be part of an exact series expansion of the certainty equivalent. For example, consider the risk-neutral case $c(\boldsymbol{a}, \mu) = -\beta^{-1}\log \mathrm{E}_\mu e^{-\beta \mathrm{E}_m \boldsymbol{a}}$ in which $u(x) = x$ and $v(x) = -e^{-\beta x}$ (Example 1). Its quadratic approximation (35) is $c(w + \boldsymbol{b}, \mu) \approx w + \mathrm{E}_{\bar{\mu}}\boldsymbol{b} - (\beta/2)\sigma_\mu^2(\mathrm{E}(\boldsymbol{b}))$. At the same time, under standard regularity conditions we have, for $\beta \neq 0$ small enough, the expansion (see the Appendix)

$$c(w + \boldsymbol{b}, \mu) = w + \mathrm{E}_{\bar{\mu}}\boldsymbol{b} - \frac{1}{2}\beta\sigma_\mu^2(\mathrm{E}(\boldsymbol{b})) + \sum_{n=3}^{\infty}\frac{(-1)^{n+1}}{n!}\beta^{n-1}c_n(\mathrm{E}(\boldsymbol{b})). \tag{40}$$

Here $c_n(\mathrm{E}(\boldsymbol{b}))$ is the nth cumulant of the random variable $\mathrm{E}(\boldsymbol{b}) : M \to \mathbb{R}$, which can be expressed in terms of higher moments (with the third and fourth cumulants related to skewness and kurtosis, respectively; see, for example, Kendall 1946, Chap. 3).

When the prior is such that the random variable $\mathrm{E}(\boldsymbol{b})$ has a normal distribution with parameters k and σ_μ^2, the cumulants $n \geq 3$ are zero and so the expansion (40) takes the quadratic form (39); that is, $c(w + \boldsymbol{b}, \mu) = w + k - (\beta/2)\sigma_\mu^2$. However, unlike (39) this quadratic expansion does not rely on any normality assumption on $\boldsymbol{b}$ itself.

4.8. Behavioral Foundation

The smooth ambiguity representation admits a simple behavioral foundation in a classical decision problem $(\boldsymbol{A}, S, C, M, \succsim)$ when M is a finite orthogonal set and acts are monetary (Section 3.2).[82]

81. The normality of the prior implies that $\bar{k}$ is the mean of the unknown means k and that σ_μ^2 is their variance.

82. This section was written in collaboration with Fabio Maccheroni. For a related axiomatization, in an intertemporal setting, see Cerreia-Vioglio et al. (2013a).

Specifically, denote by $\mathcal{E} = \{E_m\}_{m \in M}$ the event partition such that $m(E_m) = 1$ and $m(E_{m'}) = 0$ if $m' \neq m$ (Section 3.2). We make a few assumptions, most of which rely on that partition. We begin with a basic requirement on the preference.

ASSUMPTION 1. The preference $\succsim$ is both complete and transitive.

Next we assume that subjective expected utility applies to acts that are $\mathcal{E}$-measurable.

ASSUMPTION 2. There exists a continuous and strictly increasing function $v : \mathbb{R} \to \mathbb{R}$ and a probability measure P on $\mathcal{E}$ such that $\boldsymbol{a} \sim v^{-1}(\int_S v(\boldsymbol{a}(s)) dP(s))$ for each $\mathcal{E}$-measurable act $\boldsymbol{a} \in \boldsymbol{A}$.

Next we assume a famous Savage axiom, but on a restricted class of acts. Given any two acts $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{A}$ and any event E, set

$$\boldsymbol{a}E\boldsymbol{b} = \begin{cases} \boldsymbol{a}(s) & \text{if} \quad s \in E \\ \boldsymbol{b}(s) & \text{if} \quad s \notin E. \end{cases}$$

In words, $\boldsymbol{a}E\boldsymbol{b}$ is the act that "mixes" acts $\boldsymbol{a}$ and $\boldsymbol{b}$ via event E.

ASSUMPTION 3 (*Conditional Sure Thing Principle*). If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d} \in \boldsymbol{A}$ and $E \in \mathcal{E}$, then

$$\boldsymbol{a}E\boldsymbol{c} \succsim \boldsymbol{b}E\boldsymbol{c} \iff \boldsymbol{a}E\boldsymbol{d} \succsim \boldsymbol{b}E\boldsymbol{d}.$$

According to this separability requirement, the ranking of acts is independent of common parts that depend on events, such as E^c, belonging to the algebra generated by $\mathcal{E}$. This assumption allows us to define, for each $E \in \mathcal{E}$, the conditional preference $\succsim_E$ by

$$\boldsymbol{a} \succsim_E \boldsymbol{b} \iff \boldsymbol{a}E\boldsymbol{c} \succsim \boldsymbol{b}E\boldsymbol{c} \qquad \forall \boldsymbol{c} \in \boldsymbol{A}.$$

Next we assume that each such conditional preference has an expected utility representation.

ASSUMPTION 4. There exists a continuous and strictly increasing function $u : \mathbb{R} \to \mathbb{R}$ such that, for each $m \in M$, we have $\boldsymbol{a} \sim_{E_m} u^{-1}(\int_S u(\boldsymbol{a}(s)) dm(s))$ for all $\boldsymbol{a} \in A$.

We can now establish the desired behavioral foundation for smooth ambiguity.[83]

PROPOSITION 6. *A preference $\succsim$ on $\boldsymbol{A}$ satisfies Assumptions* 1–4 *if and only if there exists a probability measure μ on M such that representation* (17) *holds, that is, $V(\boldsymbol{a}) = v^{-1}(\int_M v(U(\boldsymbol{a}, m)) d\mu(m))$. In particular, $\bar{\mu} = P$ on $\mathcal{E}$.*

The probability P in Assumption 2 is nothing but the restriction on the partition $\mathcal{E}$ of the predictive probability $\bar{\mu}$.

83. Unlike Assumption 2, Assumptions 1 and 3 are not in behavioral terms; it is, however, a routine exercise to translate them into such terms through existing behavioral foundations of subjective expected utility in finite and infinite state spaces.

To understand this result, note that orthogonality entails that each model m identifies a single atom E_m of the partition $\mathcal{E}$. In fact, conditional on event E_m, only the physical risk described by m matters, that is, $\bar{\mu}(E|E_m) = m(E|E_m)$. The relative likelihood of different models thus corresponds to the relative likelihood of the different atoms. In other words, model uncertainty reduces to uncertainty over these atoms. Assumption 2 thus amounts to assuming expected utility for this kind of uncertainty. Given any atom, physical uncertainty emerges through the conditional preference $\succsim_{E_m}$; as a result, Assumption 4 assumes expected utility also for this other kind of uncertainty. The proof of Proposition 6 (see the Appendix) shows how to combine the two separate expected utility assumptions into the representation (17).

4.9. Optima

Here we consider optimization problems of the form $\max_{\boldsymbol{a}\in \boldsymbol{A}} V(\boldsymbol{a})$, which play a key role in applications.

Back to Actions. Until now we have considered a few decision criteria within Savage's framework $(\boldsymbol{A}, S, C, \succsim)$, with acts rather than actions as the objects of choice. Although this framework is the one best suited for theoretical analysis, in applications it is often more convenient to deal with actions than with acts, and so to consider the action-based classical decision problem $(A, S, C, \rho, M, \succsim)$. Actions can be easier to interpret, as remarked in Section 2.4, and can be represented as vectors; hence they can be easier to handle analytically than acts, which are functions.

Be that as it may, we refer to $r = u \circ \rho : A \times S \to \mathbb{R}$ as the *payoff* (or *reward*) function. The expected payoff $R(a,m) = \int_S r(a,s)dm(s)$ of action a is simply the expected utility $U(\boldsymbol{a}, m)$ of the associated act $\boldsymbol{a}$. The smooth criterion can then be written, in shorthand notation, as $V(a) = \mathrm{E}_\mu \phi(R(a,m)) \equiv \int_M \phi(R(a,m))d\mu(m)$.

Optima. We are now in a position to deal with the previous optimization problem, which can be written in terms of actions as

$$\max_{a\in A} \ V(a) = \max_{a\in A} \ \mathrm{E}_\mu \phi(R(a,m)). \tag{41}$$

The following proposition, which extends a portfolio result of Taboga (2005) and Gollier (2011), shows how this problem is equivalent to an expected utility problem when the original prior μ is replaced by a tilted version $\hat{\mu}$, whose form depends on all elements of the decision problem at hand, including ambiguity attitudes.

PROPOSITION 7. *Suppose that $\phi : \mathrm{Im}\, R \to \mathbb{R}$ is twice differentiable, with $\phi' > 0$ and $\phi'' < 0$. For each $s \in S$, let $r(\cdot, s) : A \subseteq \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and strictly concave on the convex set A. Then there exists a probability measure $\hat{\mu} \in \Delta(M)$, equivalent to μ, such that problems*

$$\max_{a\in A} \ \mathrm{E}_\mu \phi(R(a,m)) \quad \text{and} \quad \max_{a\in A} \ \mathrm{E}_{\hat{\mu}} R(a,m) \tag{42}$$

share the same solution $\hat{a}$*. In particular,* $\hat{\mu}(m) = \zeta(m)\mu(m)$ *where* $\zeta : M \to \mathbb{R}$ *is the Radon–Nikodym function*

$$\zeta(m) = \frac{\phi'(R(\hat{a}, m))}{\mathrm{E}_{\mu}\phi'(R(\hat{a}, m))}. \tag{43}$$

Since ϕ' is decreasing, the tilted prior $\hat{\mu}$ alters, through ζ, the original prior μ by shifting weight to models m with a lower expected utility $R(\hat{a}, m)$, thus magnifying their relative importance. Formally,

$$\zeta(m') < \zeta(m) \iff R(\hat{a}, m) < R(\hat{a}, m').$$

In sum, action $\hat{a}$ solves problem $\max_{a \in A} \mathrm{E}_{\hat{\mu}} R(a, m)$ even though $\hat{\mu}$ handicaps $\hat{a}$ by overweighting its cons—at the expense of its pros—relative to the original prior μ. In this sense, $\hat{a}$ is a robust solution when compared to the solution of the expected utility problem $\max_{a \in A} \mathrm{E}_{\mu} R(a, m)$. Ambiguity aversion can thus be seen as a desire for robustness on the problem solutions.[84]

EXAMPLE 2. In the CARA case $\phi(x) = -e^{-\lambda x}$, we have the exponential tilt $\zeta(m) \propto e^{-\lambda R(\hat{a}, m)}$. In the CRRA case $\phi(x) = x^{\lambda}$, we have the power tilt $\zeta(m) \propto R(\hat{a}, m)^{\lambda - 1}$. ■

An action $a \in A$ is crisp for a smooth preference $\succsim$ if $R(a, m) = R(a, m')$ for all $m, m' \in \operatorname{supp} \mu$ (see Section 3.1). Crisp actions are not sensitive to model uncertainty and so for them $\zeta = 1$; this equality implies the following corollary.

COROLLARY 1. *Under the hypotheses of Proposition 7, the two problems in* (41) *share the same solution if it is crisp.*

In short, ambiguity attitudes may matter only for decision problems that do not have crisp solutions.

We close by observing that the tilted prior can be also regarded as a mixture, with weight determined by ξ, of the original prior and of another prior. Tilting can be thus seen as a convexification of priors.

LEMMA 2. *There exist* $\alpha \in (0, 1]$ *and* $\nu \in \Delta(M)$ *such that* $\hat{\mu} = \alpha\mu + (1 - \alpha)\nu$ *if and only if* $\alpha \leq \min_{m \in \operatorname{supp} \mu} \zeta(m)$.[85]

Monetary Policy Illustration. To illustrate the previous result, we continue our study of the monetary policy problem (Section 3.1). For that purpose we define a bivariate

84. For instance, in environmental policy making this feature of ambiguity aversion can provide a rationale for some forms of the precautionary principle (Stern 2007). The more stringent the principle, the higher the underlying ambiguity aversion (until the limit maxmin criteria of Section 4.6).

85. When M is not finite, the lemma requires the (almost everywhere) boundedness of ζ.

random variable $(\boldsymbol{u}, \boldsymbol{\pi})$ by

$$\boldsymbol{u}(a, w, \varepsilon, \theta) = \theta_0 + (\theta_{1\pi} + \theta_{1a})a + \theta_{1\pi}\theta_3\varepsilon + \theta_2 w$$
$$\boldsymbol{\pi}(a, w, \varepsilon, \theta) = a + \theta_3\varepsilon.$$

This allows us to rewrite the outcome function as $\rho(a, w, \varepsilon, \theta) = (\boldsymbol{u}(a, w, \varepsilon, \theta), \boldsymbol{\pi}(a, w, \varepsilon, \theta))$. We make two assumptions:

(i) shocks ε and w are uncorrelated and have zero mean and unit variance with respect to the known distribution q;

(ii) the policy multiplier is negative, that is, $\theta_{1\pi} + \theta_{1a} \leq 0$.

Consider the quadratic payoff case $r = -\boldsymbol{u}^2 - \boldsymbol{\pi}^2$. That puts us in the classic linear quadratic policy framework à la Tinbergen (1952) and Theil (1961), with objective function

$$\begin{aligned} V(a) &= \mathrm{E}_\mu \phi(\mathrm{E}_q(-\boldsymbol{u}^2 - \boldsymbol{\pi}^2)) \\ &\equiv \int_\Theta \phi\left(-\int \boldsymbol{u}^2(a, w, \varepsilon, \theta)dq - \int \boldsymbol{\pi}^2(a, w, \varepsilon, \theta)dq\right) d\mu(\theta) \\ &= \int_\Theta \phi\left(-(\theta_0 + (\theta_{1\pi} + \theta_{1a})a)^2 - a^2 - \theta_2^2 - \theta_3^2\theta_{1\pi}^2 - \theta_3^2\right) d\mu(\theta). \end{aligned}$$

Let θ^* be the true model economy. To ease the analysis, assume that the monetary authority knows the true values $\theta^*_{1\pi}$, θ^*_2 and θ^*_3, that is, the shocks' coefficients and the slope of the Phillips curve.[86] The objective function can then be written as

$$V(a) = \mathrm{E}_\mu \phi(-(\theta_0 + (\theta_{1\pi} + \theta_{1a})a)^2 - a^2 + \text{const.})$$

where now $\theta = (\theta_0, \theta_{1a}) \in \Theta$. When the monetary authority knows the true model economy, and therefore faces only the physical risk of shocks, the (objectively) optimal policy is

$$a^o = B(\theta^*) = -\frac{\theta_0^*\left(\theta^*_{1\pi} + \theta^*_{1a}\right)}{1 + \left(\theta^*_{1\pi} + \theta^*_{1a}\right)^2}$$

where $B : \Theta \to A$ is the best reply function. In contrast, when the true model economy is unknown, some simple algebra shows that the optimal policy becomes

$$\hat{a} = B(\hat{\mu}) = -\frac{\mathrm{E}_{\hat{\mu}}(\theta_0)\big(\theta^*_{1\pi} + \mathrm{E}_{\hat{\mu}}(\theta_{1a})\big) + \mathrm{Cov}_{\hat{\mu}}(\theta_0, \theta_{1a})}{1 + \big(\theta^*_{1\pi} + \mathrm{E}_{\hat{\mu}}(\theta_{1a}) big)^2 + V_{\hat{\mu}}(\theta_{1a})},$$

where $B : \Delta(\Theta) \to A$ is the expected utility best reply function with respect to the tilted prior $\hat{\mu}$ (which reduces to μ when ϕ is linear).

86. Battigalli et al. (2015b) show that this assumption holds in steady state.

Action $B(\hat{\mu})$ is the robust version of action $B(\mu)$ that takes ambiguity aversion into account. More robust does not necessarily mean more prudent, that is, $B(\hat{\mu}) \leq B(\mu)$. For instance, suppose the monetary authority has a dogmatic prior about the coefficient θ_{1a}, that is, there is some value $\bar{\theta}_{1a}$ such that $\mu(\bar{\theta}_{1a}) = 1$. For example, $\bar{\theta}_{1a} = 0$ when the authority believes that the true model economy is of the Lucas–Sargent type (with a zero policy multiplier), while $\bar{\theta}_{1a} = -\theta^*_{1\pi}$ when the authority believes that the true model is of the Samuelson–Solow type (with a nonzero policy multiplier). In any case, being the prior equivalent to its tilted version, we also have $\hat{\mu}(\bar{\theta}_{1a}) = 1$. Therefore,[87]

$$B(\hat{\mu}) \leq B(\mu) \Longleftrightarrow \mathrm{E}_{\hat{\mu}}(\theta_0) \leq \mathrm{E}_{\mu}(\theta_0).$$

The robust policy is less aggressive as long as the tilted expected value of coefficient θ_0 is lower than the original one. Note that for the monetary authority that believes in the Lucas–Sargent model economy we have $B(\mu) = B(\hat{\mu}) = 0$; in this case, a zero-target-inflation policy is optimal irrespective of the uncertainty about the coefficient θ_0.

4.10. Illustration: Static Asset Pricing

We consider a two-period economy populated by agents who play the dual role of consumers and investors.[88] As consumers, they allocate consumption across the two periods; as investors they fund such allocation by trading assets in a financial market. Our purpose here is examining how market asset prices relate to agents' individual consumption allocations when those agents are ambiguity averse.

Financial Market. In a two-period frictionless financial market, at date 0 (today) investors trade n primary assets—in any quantity and without any kind of impediment (transaction costs, short sales constraints, etc.)—that pay out at date 1 (tomorrow) contingent on which state $s \in S = \{s_1, \ldots, s_k\}$ obtains. The true probability model on S is m^*, that is, $m^*(s)$ is the probability that state s obtains.

To ease matters, we assume that: (i) there is a single consumption good in the economy (potatoes) and the assets payoffs are in terms of this good; (ii) all states are essential, that is, $m^*(s) > 0$ for each $s \in S$.

Let $L = \{y_1, \ldots, y_n\} \subseteq \mathbb{R}^k$ be the collection of primary assets and let $p = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n$ be the vector of their market prices (in terms of potatoes). The triple (L, p, m^*) describes the financial market.

Portfolios and Contingent Claims. A primary asset $j = 1, \ldots, n$ is denoted by $y_j = (y_{1j}, \ldots, y_{kj}) \in \mathbb{R}^k$, where y_{ij} represents its payoff if state s_i obtains.

87. In fact, $V_\mu(\theta_{1a}) = \mathrm{Cov}_\mu(\theta_0, \theta_{1a}) = V_{\hat{\mu}}(\theta_{1a}) = \mathrm{Cov}_{\hat{\mu}}(\theta_0, \theta_{1a}) = 0$. Note that under dogmatic beliefs on θ_{1a} the certainty equivalence principle holds for $B(\mu)$. Actually, under the condition $\mathbb{E}_\mu(\theta_0)(\theta^*_{1\pi} + \mathbb{E}_\mu(\theta_{1a}))V_\mu(\theta_{1a}) \neq \mathrm{Cov}_\mu(\theta_0, \theta_{1a}) + (\theta^*_{1\pi} + \mathbb{E}_\mu(\theta_{1a}))^2$, this is the only case when this holds.

88. This section was written in collaboration with Nicola Rosaia.

Portfolios of primary assets can be formed in the market, each identified by a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ where x_j is the traded quantity of primary asset y_j.[89] In particular, the primary asset y_1 is identified by the portfolio $e^1 = (1, 0, \dots, 0) \in \mathbb{R}^n$, the primary asset y_2 by $e^2 = (0, 1, 0 \dots, 0) \in \mathbb{R}^n$, and so on. The linear combination $x \cdot y = \sum_{j=1}^n x_j y_j \in \mathbb{R}^k$ is the state-contingent payoff that, tomorrow, portfolio x ensures.

We call any state-contingent payoff $w \in \mathbb{R}^k$ a *contingent claim*. A claim w is *replicable* (in the market) if there exists a portfolio x such that $w = \sum_{j=1}^n x_j y_j$. In words, replicable contingent claims are the state-contingent payoffs that, tomorrow, can be attained by trading, today, primary assets. The *market* W is the vector subspace of $\mathbb{R}^k$ consisting of all replicable contingent claims; that is, $W = \operatorname{span} L$.

The market is *complete* if $W = \mathbb{R}^k$: if so, all contingent claims are replicable. Otherwise, the market is *incomplete*. By a basic linear algebra result, completeness of the market amounts to the replicability of the k *Arrow* (or *pure*) contingent claims $e^i \in \mathbb{R}^k$ that pay out 1 euro if state s_i obtains and 0 otherwise. These important claims uniquely identify states.

Market Value. We can represent the collection L of primary assets by the payoff matrix

$$Y = (y_{ij}) = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ y_{k1} & y_{k2} & \cdots & y_{kn} \end{bmatrix}$$

which has k rows (states) and n columns (assets); entry y_{ij} represents the payoff of primary asset y_j in state s_i. The linear operator $R : \mathbb{R}^n \to \mathbb{R}^k$ given by $R(x) = Yx$ describes the contingent claim determined by portfolio x. In other words, $R_i(x)$ is the payoff of portfolio x if state s_i obtains. Clearly, $W = \operatorname{Im} R$ and so the rank of the payoff matrix Y is the dimension of the market W.

In a frictionless market, the (market) *value* $p \cdot x = \sum_{j=1}^n p_j x_j$ of a portfolio x is the cost of the market operations it requires. The (market) *value function* $v : \mathbb{R}^n \to \mathbb{R}$ is the linear function that assigns to each portfolio x its value $v(x)$. Note that it is the market's frictionless nature that ensures the value function's linearity.

The value of primary assets is their price. For, recalling that the primary asset y_j is identified by the portfolio e^j, we have

$$v(e^j) = p \cdot e^j = p_j. \tag{44}$$

89. If $x_j \geq 0$ (resp., $x_j \leq 0$) the portfolio is long (resp., short) on asset y_j, that is, it buys (resp., sells) x_j units of the asset.

Law of One Price. The *law of one price* (LOP) is a fundamental property of a financial market (L, p, m^*): for all portfolios $x, x' \in \mathbb{R}^n$,

$$R(x) = R(x') \Longrightarrow v(x) = v(x'). \tag{45}$$

In words, portfolios that induce the same contingent claims must share the same market value.[90] In fact, the contingent claims that they determine is all that matters in portfolios, which are simply instruments to achieve those claims. If two portfolios inducing the same contingent claim had different market values, then a (sure) saving opportunity would be missed in the market. The LOP requires the financial market to take advantage of any such opportunity.

The financial market satisfies the LOP if and only if the set $\{v(x) : x \in R^{-1}(w)\}$ is a singleton for each $w \in W$.[91] So, under the LOP all portfolios x that replicate a contingent claim w share the same value $v(x)$. It is natural to regard such common value as the price of the claim: we define the *price* p_w of a replicable contingent claim $w \in W$ to be the value of a replicating portfolio x; that is, $p_w = v(x)$ where $w = R(x)$. In words, p_w is the market cost $v(x)$ incurred today to form a portfolio x that tomorrow will ensure the contingent claim w, that is, $w = R(x)$. For primary assets we are back to (44), with $p_j = v(e^j)$.

Stochastic Discount Factors. The *pricing rule* $f : W \to \mathbb{R}$ associates with each replicable contingent claim $w \in W$ its price p_w—that is, $f(w) = p_w$. It is easy to check that f inherits the linearity of the market value function v.[92] Thanks to the LOP, the pricing rule is thus a well-defined linear function. This simple but key observation is part of the so-called fundamental theorem of asset pricing.

Since f is linear, by a basic linear algebra result there exists a unique vector $\rho \in W$, the *stochastic discount factor,* such that

$$f(w) = \mathrm{E}_{m^*}(\rho w) \qquad \forall w \in W. \tag{46}$$

Since it belongs to W, the stochastic discount factor itself can be regarded as a contingent claim (see Hansen and Richard 1987).

Two brief remarks may shed light on the nature of the stochastic discount factor. First, in a complete market we have $\rho_i = p_{e^i}/m_i^*$, where p_{e^i} is the price of the Arrow contingent claim e^i; in words, the stochastic discount factor is the vector price of the Arrow securities, normalized by the state probabilities. Second, if the constant (and so risk free) contingent claim $\mathbf{1} = (1, \ldots, 1)$ is replicable, as when the market is complete, then $\mathrm{E}_{m^*}(\rho) = p_{\mathbf{1}}$. The expected value of the stochastic discount factor is

90. If not all states were essential, the equality $R(x) = R(x')$ would be required only to hold m^*-almost everywhere (i.e., only for essential states).

91. Note that $x \in R^{-1}(w)$ if x replicates w.

92. Nonlinearities of the market value function, due to market imperfections, would translate into nonlinearities of the pricing rule—regardless of whether or not the market is complete (see Cerreia-Vioglio, Maccheroni, and Marinacci 2015, and references therein).

the price of such risk-free claim. Equivalently, $r_f = 1/\mathrm{E}_{m^*}(\rho)$ where r_f is the (gross) return of the risk-free asset, that is, the reciprocal of its price.

Consumers. In our two-period economy, agents in their role of consumers have to decide how much to consume today, c_0, and tomorrow, c_1, of the economy single consumption good (potatoes). Today's consumption c_0 is a positive scalar, whereas tomorrow's consumption is state contingent and therefore a vector $c_1 \in \mathbb{R}^k_+$. As a consumer, the agent's object of choice is thus a contingent good $c = (c_0, c_1) \in \mathbb{R}_+ \times \mathbb{R}^k_+$.

As "classical" DMs (Section 2.5), agents posit a collection M of models that contain the true model m^*. To facilitate the analysis we assume that for, all posited models $m \in M$, the inequality $m(s) > 0$ holds for all $s \in S$. That is, all models classify all states as essential.[93]

Agents rank contingent goods according to the smooth objective function

$$V(c) = \mathrm{E}_\mu \phi(\mathrm{E}_m u(c_0, c_1)), \tag{47}$$

where u is a von Neumann–Morgenstern intertemporal utility function (monotonic increasing in its arguments), ϕ models ambiguity attitudes, and μ is a prior probability.

Budget Set. Agents have endowment profiles $I = (I_0, I_1) \in \mathbb{R}_+ \times \mathbb{R}^k_+$, where I_0 is today's endowment and I_1 is tomorrow's state-contingent endowment. They can fund their consumption by using their endowment and, in their investor role, by trading assets. Since u is monotonic, their (relevant) budget set is

$$B(p, I) = \left\{ (c_0, c_1, x) : c_0 + p \cdot x = I_0 \text{ and } c_{1i} = I_{1i} + \sum_{j=1}^{n} x_j y_{ij} \, \forall i \right\}.$$

In the language of contingent claims, we can equivalently write

$$B(p, I) = \left\{ (c_0, c_1, w) : c_0 + p_w = I_0 \text{ and } c_{1i} = I_{1i} + w_i \, \forall i \right\}$$

where $w \in W$ is any replicable contingent claim, with price p_w. Under the LOP, by (46) we can then write

$$B(p, I) = \left\{ (c_0, c_1, w) : c_0 + \mathrm{E}_{m^*}(\rho w) = I_0 \text{ and } c_{1i} = I_{1i} + w_i \, \forall i \right\}, \tag{48}$$

where ρ is the stochastic discount factor.

93. What matters in this assumption is that it makes all models equivalent—that is, they agree on which events have probability 0 or 1. Hence, all equalities that, like (45) and (48), involve state-contingent quantities can be required to hold almost everywhere with respect to any model. The analysis would still hold under this equivalence assumption on models, which is a weaker assumption than the essentialness of all states.

Consumption/Investment Problem. Because agents in the economy are both consumers and investors, they must decide both their consumption and asset allocations. Given the LOP, they solve the optimization problem

$$\max_{c,w} \; V(c) \quad \text{sub } (c, w) \in B(p, I), \tag{49}$$

where the objective function is given by (47) and the budget constraint by (48).

Valuation. Next we state a simple necessary condition for the previous optimization problem.

PROPOSITION 8. *Suppose that the function ϕ is twice differentiable, with $\phi' > 0$ and $\phi'' < 0$. Let u be strictly concave and differentiable, with strictly positive gradient and $\partial u/\partial c_0$ independent of c_1. Then there exists at most an interior solution $\hat{c} \in \mathbb{R}_+^{k+1}$ of problem* (48)*, and it satisfies the condition*

$$p_w = \mathrm{E}_\mu \left(\frac{\phi'(\mathrm{E}_m u(\hat{c}))}{\mathrm{E}_\mu \phi'(\mathrm{E}_m u(\hat{c}))} \mathrm{E}_m \left(\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})} w \right) \right) \qquad \forall w \in W. \tag{50}$$

This formula relates market prices of assets and agents' individual consumption decisions in the two-period economy. It shows that both ambiguity and risk attitudes affect asset pricing. If the market is complete, it implies

$$\rho_i = \frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})} \frac{\mathrm{E}_\mu \left(\phi'(\mathrm{E}_m u(\hat{c}))(m_i/m_i^*)\right)}{\mathrm{E}_\mu \phi'(\mathrm{E}_m u(\hat{c}))} \qquad \forall i = 1, \ldots, k. \tag{51}$$

When the true model m^* is known, (50) reduces to the usual (see, e.g., Cochrane and Hansen 1992; Cochrane 2005) pricing formula

$$p_w = \mathrm{E}_{m^*} \left(\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})} w \right), \tag{52}$$

which is based on the intertemporal rate of substitution in consumption. In general, (50) implies that prices can be decomposed as

$$\begin{aligned} p_w &= \mathrm{E}_{\bar{\mu}} \left(\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})} w \right) + \mathrm{Cov}_\mu \left(\frac{\phi'(\mathrm{E}_m u(\hat{c}))}{\mathrm{E}_\mu \phi'(\mathrm{E}_m u(\hat{c}))}, \mathrm{E}_m \left(\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})} w \right) \right) \\ &\equiv q_w + \mathrm{Cov}_\mu \left(\zeta(m), p_w^m \right). \end{aligned}$$

Here q_w is the ambiguity neutral price of claim w that corresponds to $\hat{c}$ and p_w^m is the rational expectations price of claim w that corresponds to $\hat{c}$ under model m—that is, the price of the claim if that model is known to be true (or, at least, if $\bar{\mu} = m^*$).

The difference $p_w - q_w$ can be viewed as an ambiguity adjustment. Let m_g be a model that, if known to be true, has a higher subjective value than m_b—that is,

$\mathrm{E}_{m_g} u(\hat{c}) > \mathrm{E}_{m_b} u(\hat{c})$. This amounts to $\zeta(m_g) < \zeta(m_b)$. We have $\mathrm{Cov}_\mu(\zeta(m), p_w^m) \leq 0$ if this implies that m_g has a higher rational expectations market value than does m_b—namely, $p_w^{m_g} > p_w^{m_b}$. Thus, the price of a claim incorporates a negative ambiguity adjustment when models that, if known to be true, have a higher subjective value give the claim a higher rational expectations price.

EXAMPLE 3. In the time separable case $u(c_0, c_1) = \mathsf{v}(c_0) + \beta \mathsf{v}(c_1)$, we have

$$p_w = \beta \mathrm{E}_\mu \left(\frac{\phi'(\mathsf{v}(c_0) + \beta \mathrm{E}_m \mathsf{v}(\hat{c}_1))}{\mathrm{E}_\mu(\phi'(\mathsf{v}(c_0) + \beta \mathrm{E}_m \mathsf{v}(\hat{c}_1)))} \mathrm{E}_m \left(\frac{\mathsf{v}'(\hat{c}_1)}{\mathsf{v}'(\hat{c}_0)} w \right) \right).$$

In the CRRA case $\mathsf{v}(x) = x^\gamma$, for CRRA and CARA forms for ϕ (see Section 4.2) we have:

(i) if $\phi(x) = x^\theta$, then

$$p_w = \beta \mathrm{E}_\mu \left(\frac{\left(\hat{c}_0^\gamma + \beta \mathrm{E}_m \hat{c}_1^\gamma\right)^{\theta-1}}{\mathrm{E}_\mu \left(\hat{c}_0^\gamma + \beta \mathrm{E}_m \hat{c}_1^\gamma\right)^{\theta-1}} \mathrm{E}_m \left(\left(\frac{\hat{c}_1}{\hat{c}_0} \right)^{\gamma-1} w \right) \right);$$

(ii) if $\phi(x) = -e^{-\theta x}$, then

$$p_w = \beta \mathrm{E}_\mu \left(\frac{e^{-\beta\theta \mathrm{E}_m \hat{c}_1^\gamma}}{\mathrm{E}_\mu e^{-\beta\theta \mathrm{E}_m \hat{c}_1^\gamma}} \mathrm{E}_m \left(\left(\frac{\hat{c}_1}{\hat{c}_0} \right)^{\gamma-1} w \right) \right). \qquad \blacksquare$$

As remarked previously, the connection established by formula (50) has been the focus of this section. Similar pricing formulas can be found in a series of papers by Lars Peter Hansen and Thomas Sargent (e.g., Hansen 2007; Hansen and Sargent 2007, 2008, 2010, 2014), which also study their empirical relevance in explaining some asset pricing phenomena that the standard formula (52) is unable to explain (unless contrived risk aversion assumptions are made on the function u). Related formulas are investigated by Hayashi and Miao (2011), Collard et al. (2012), Ju and Miao (2012), Jahan-Parvar and Liu (2014), Backus, Ferriere, and Zin (2015), and Guerdjikova and Sciubba (2015).

Neutral Valuation. Under physical uncertainty, risk-neutral probabilities play a key role in asset pricing. In our two-stage setup, with both physical and model uncertainty, we can establish an expected value form—neutral with respect to uncertainty attitudes—of the asset pricing formula (50). To this end, define $\hat{m} \in \Delta(S)$ by

$$\hat{m}_i = \xi_i^m m_i \qquad \forall i = 1, \ldots, k$$

and, as in Proposition 7, $\hat{\mu} \in \Delta(M)$ by $\hat{\mu}(m) = \zeta(m)\mu(m)$ for every $m \in M$, where $\xi^m : S \to \mathbb{R}$ and $\zeta : M \to \mathbb{R}$ are, respectively, the Radon–Nikodym functions given by

$$\xi_i^m = \frac{\partial u}{\partial c_{1i}}(\hat{c}) / \mathrm{E}_m \frac{\partial u}{\partial c_1}(\hat{c})$$

and

$$\zeta(m) = \phi'(\mathrm{E}_m u(\hat{c}))/\mathrm{E}_\mu \phi'(\mathrm{E}_m u(\hat{c})).^{94}$$

The tilted model $\hat{m}$ is the (equivalent) risk-neutral version of the model m, and the tilted prior $\hat{\mu}$ is the (equivalent) ambiguity neutral version of the prior μ. In particular, $\hat{m} = m$ under risk neutrality (u linear) and $\hat{\mu} = \mu$ under ambiguity neutrality (ϕ linear). Next, define $\tilde{\mu} \in \Delta(M)$ by

$$\tilde{\mu}(m) = \chi(m)\hat{\mu}(m) \qquad \forall m \in M,$$

where $\chi : M \to \mathbb{R}$ is the Radon–Nikodym function given by

$$\chi(m) = \mathrm{E}_m \frac{\partial u}{\partial c_1}(\hat{c})/\mathrm{E}_{\hat{\mu}} \mathrm{E}_m \frac{\partial u}{\partial c_1}(\hat{c}).$$

In particular, $\tilde{\mu} = \hat{\mu}$ if χ is crisp: $\chi(m) = \chi(m')$ for all $m, m' \in \operatorname{supp} \mu$, that is, the expected marginal utility of tomorrow consumption

$$\mathrm{E}_m \frac{\partial u}{\partial c_1}(\hat{c})$$

is unaffected by model uncertainty.[95] Under ambiguity neutrality we may well have $\tilde{\mu} \neq \hat{\mu} = \mu$ if χ is not crisp, that is, if model uncertainty matters.

PROPOSITION 9. *Suppose the risk-free claim* **1** *is replicable. Under the hypotheses of Proposition* 8*,*

$$p_w = \frac{1}{r_f} \mathrm{E}_{\tilde{\mu}}(\mathrm{E}_{\hat{m}} w) \quad \forall w \in W. \tag{53}$$

The pricing formula (53) is the expected value version of (50). The tilted predictive probability $\overline{\tilde{\mu}} \in \Delta(S)$ given by $\overline{\tilde{\mu}} = \mathrm{E}_{\tilde{\mu}} \hat{m}$ allows us to write it more compactly as

$$p_w = \frac{1}{r_f} \mathrm{E}_{\overline{\tilde{\mu}}}(w). \tag{54}$$

We can thus regard $\overline{\tilde{\mu}}$ as the uncertainty neutral measure that asset pricing with model uncertainty features. When the true model m^* is known, (54) reduces to the usual (see, e.g., Ross 2005, p. 8) risk-neutral valuation $p_w = \mathrm{E}_{\hat{m}^*}(w)/r_f$.

94. Vectors in $\mathbb{R}^k$ can be equivalently seen as real valued functions on S; hence, we interchangeably use (according to convenience) the vector and function notations $\xi^m \in \mathbb{R}^k$ and $\xi^m : S \to \mathbb{R}$.

95. Consumption data can be used to estimate ζ, χ, and ξ^m (if agents are assumed to act optimally; see Cochrane and Hansen 1992).

4.11. Diversification

Preference for action diversification is a basic economic principle. In choice under certainty, it reflects DMs' desire for variety. In choice under uncertainty, it reflects their desire to hedge against uncertainty in order "not to put all eggs in one basket". Ambiguity aversion, with its desire for robustness, magnifies such preference.

To illustrate this important issue in the present two-stage setting, we consider the policy problem of public officials needing to decide which treatment should be administered to a population (Section 2.2). Owing to the treatment diversification that they permit, fractional treatment actions may be a relevant option because, under state uncertainty, individuals with the same covariate may respond differently to the same treatment. As emphasized by Manski (2009), standard expected utility is unable to give fractional treatments the relevance that, intuitively, they seem to deserve. For this reason, he replaced expected utility with maxmin regret, a criterion due to Savage (1951) that, however, violates the independence of irrelevant alternatives, a basic rationality tenet.[96] A proper account of model uncertainty may, however, justify fractional treatment without having to invoke a departure from expected utility as radical as maxmin regret.

To see why this is the case, for simplicity we assume that the population is homogeneous (with all individuals sharing the same covariate). In this case, a treatment action is a probability distribution $a \in A = \Delta(T)$ over the finite collection T of alternative treatments, where $a(t)$ is the fraction of the population that has been assigned treatment t. The consequence function is $\rho(a, s) = \sum_{t \in T} c(t, s) a(t)$. Denote by $\bar{c}_m(t) = \sum_{s \in S} c(t, s) m(s)$ the expected outcome of treatment t under model m. Then criterion (19) takes the form

$$V(a) = \sum_{m \in M} \phi\left(\sum_{s \in S} \rho(a, s) m(s)\right) d\mu(m) = \sum_{m \in M} \phi\left(\sum_{t \in T} \bar{c}_m(t) a(t)\right) d\mu(m).$$

This expression is a utilitarian social welfare criterion when $c(t, s)$ is interpreted in welfare terms; when $c(t, s)$ is interpreted in material terms, it assumes that public officials are risk neutral.

Consider binary treatment problems $T = \{t_0, t_1\}$, so that we can identify actions with points of the unit interval—that is, $a \in [0, 1]$. In particular, a is the fraction of the population under treatment t_1. A treatment action is fractional when $a \in (0, 1)$—that is, when it does not assign the same treatment to all individuals. The policy problem is $\max_a V(a) = \max_a \mathrm{E}_\mu \phi((1 - a)\bar{c}_m(t_0) + a\bar{c}_m(t_1))$. When ϕ is linear, fractional treatment is not optimal unless $\bar{c}_{\bar{\mu}}(t_0) = \bar{c}_{\bar{\mu}}(t_1)$. In fact, if $\bar{c}_{\bar{\mu}}(t_1) \neq \bar{c}_{\bar{\mu}}(t_0)$, then all individuals are assigned to the treatment with the highest $\bar{c}_{\bar{\mu}}$ value.[97] As a result, in

96. See Chernoff (1954, pp. 425–426). Klibanoff (2013) presents a treatment choice example that illustrates such a failure.

97. When $\bar{c}_{\bar{\mu}}(t_0) = \bar{c}_{\bar{\mu}}(t_1)$, all actions (fractional or not) are optimal.

general there is no treatment diversification when ϕ is linear; this is the counterintuitive feature of standard expected utility just discussed.

LEMMA 3. *Suppose that ϕ is concave. Fractional treatment is suboptimal if and only if V is strictly monotonic on* $[0, 1]$.

Fractional treatment is thus suboptimal only if V is either steadily increasing or steadily decreasing in a, as when ϕ is linear and $\bar{c}_{\bar{\mu}}(t_1) \neq \bar{c}_{\bar{\mu}}(t_0)$. The next example shows that, in our setting, fractional treatment may be optimal.

EXAMPLE 4. Suppose that the outcome of interest is $c(t,s) = o(t,s) - \max_s o(t,s)$, where $o(t,s)$ is an underlying outcome that enters the objective function via the anticipated ex-post regret caused by treatment t. Since $c \leq 0$, we can consider a quadratic $\phi(x) = -x^2$. Let $d_m = \bar{c}_m(t_0) - \bar{c}_m(t_1)$. Simple algebra shows that

$$\hat{a} = \frac{\mathrm{E}_\mu \mathrm{E}_m \bar{c}_m(t_0) d_m}{\mathrm{E}_\mu d_m^2}.$$

Thus fractional treatment is optimal—that is, $\hat{a} \in (0,1)$—if and only if $|V(0) - V(1)| < \mathrm{E}_\mu d_m^2$. ■

In sum, fractional treatment may emerge under ambiguity aversion, as already remarked by Klibanoff (2013) with a related example. In general, under ambiguity aversion (ϕ concave), hedging against model uncertainty provides a further motif for action diversification, on top of the standard hedging motif against state uncertainty that risk aversion (u concave) features. In fact, suppose that A is a convex subset of some vector space.[98] To see in its purest form the diversification effect of ambiguity aversion, assume that r is linear and ϕ concave. Consider the preference $\succsim$ on A represented by the criterion $V(a) = \sum_m \phi(R(a,m))\mu(m)$. This criterion is clearly concave on A (strictly concave if ϕ is) and so, for each $\alpha \in (0,1)$, we have

$$a \sim b \Longrightarrow \alpha a + (1-\alpha) b \succsim a \tag{55}$$

with strict preference when ϕ is strictly concave. In words, if the DM is indifferent between two actions a and b, he prefers to diversify through actions that combine them. Ambiguity aversion thus features a preference for hedging, as first remarked by Schmeidler (1989), who referred to this property as *uncertainty aversion*.

Note that the abstract notion of diversification may take different meanings in different applications. For instance, consider convex sets of the form $A = \Delta(X)$. In some applications, X is a collection of "pure" actions and $a(x)$ is the probability with which a random device selects the pure action x; diversification thus amounts to randomization. In other applications, X is a collection of basic alternatives (assets, contingent goods, treatments, etc.) and $a(x)$ is the proportion of alternative x that

98. For instance, A may be a convex subset of $\mathbb{R}^n$ or it may have the form $\Delta(X)$ for some set X. In the latter case, X is embedded in A by identifying each $x \in X$ with the degenerate Dirac probability $\delta_x \in \Delta(X)$.

action a features; here diversification amounts to fractional allocation (i.e., allocation expressed in proportional terms).

5. Beyond the Bayesian Paradigm: Multiple Priors

5.1. Representation

The smooth ambiguity model drops the source independence assumption of SEU and builds on the distinction between physical and epistemic uncertainties within a Bayesian framework, in which both kinds of uncertainties are modeled via standard probability measures. But different degrees of aversion to such uncertainties are permitted, and so tastes take care of the distinction. A different, belief-based, approach that departs from the Bayesian framework originates in the seminal work of Gilboa and Schmeidler (1989) and Schmeidler (1989). Here we focus on the celebrated *multiple priors* model axiomatized by Gilboa and Schmeidler (1989).[99]

In the two-stage setup of our paper, the multiple priors model relaxes the Bayesian tenet that the DM's information about epistemic uncertainty must be quantified by a single probability measure μ and allows, instead, that it may be quantified by a set C of probability measures. Specifically, in our statistical decision theory setup based on datum M, we consider a compact set $\mathsf{C} \subseteq \Delta(M)$ of prior probabilities $\mu : 2^{M} \to [0, 1]$; this set is possibly nonsingleton because the DM may not have enough information to specify a single prior, a factor that becomes especially relevant when subjective probabilities' domains (M for μ and S for $\bar{\mu}$) are complex. The DM uses the criterion

$$W(\boldsymbol{a}) = \min_{\mu \in \mathsf{C}} \int_M \left(\int_S u(\boldsymbol{a}(s)) dm(s) \right) d\mu(m) \tag{56}$$

$$= \min_{\mu \in \mathsf{C}} \int_S u(\boldsymbol{a}(s)) d\bar{\mu}(s), \tag{57}$$

where he cautiously considers the least among all the classical SEU functionals determined by each prior μ in C. The predictive form (57) is the original version proposed by Gilboa and Schmeidler (1989), whereas (56) is the version studied by Cerreia-Vioglio et al. (2013a); both have been adapted to our setup. Behaviorally, the representation is based on a complete, transitive, and statewise monotonic preference $\succsim$ that satisfies the uncertainty aversion property (55) and a weak form of independence, as discussed in Gilboa and Marinacci (2013). In particular, the cautious attitude of this

99. This important strand of research, pioneered by David Schmeidler, is recently discussed in detail by Gilboa and Marinacci (2013). For this reason, here we just outline its basic features (for a different but closely related literature, see Wakker 2010). The relevant version of the multiple priors model for the two-stage setup that we employ is discussed by Marinacci (2002) and developed and axiomatized by Cerreia-Vioglio et al. (2013a), who relate it to robust Bayesian analysis (see Berger 1990, 1993). Other two-stage perspectives have been offered by Amarante (2009) and Giraud (2014).

maxmin criterion is based on uncertainty aversion.[100] Omitting this property yields a more general α-maxmin criterion à la Hurwicz (1951) that combines maxmin and maxmax behavior, as shown by Ghirardato, Maccheroni, and Marinacci (2004).

The uncertainty aversion property that underlies criterion (56) captures an attitude toward uncertainty that is negative but not extremely so. Criterion (56) is, indeed, less extreme than it may appear at a first sight (see Gilboa and Marinacci 2013). In fact, the set C incorporates both the attitude toward ambiguity, a taste component, and its perception, an information component. A smaller set C may reflect better information (i.e., less perceived ambiguity) and/or less aversion to uncertainty. In other words, the size of C does not reflect just information, but taste as well.[101]

In this regard, note that with singletons $\mathsf{C} = \{\mu\}$ we return to the classical SEU criterion (5). In contrast, when $\mathsf{C} = \Delta(M)$—that is, when all prior probabilities on M belong to C—we have

$$\min_{\mu \in \Delta(M)} \int_M \left(\int_S u(\boldsymbol{a}(s)) dm(s) \right) d\mu(m) = \min_{m \in M} \int_S u(\boldsymbol{a}(s)) dm(s) \tag{58}$$

that is, we return to the Waldean criterion (28). Therefore, in our setup the multiple priors criterion (56) should not be confused with the Wald maxmin criterion (28); in fact, the former provides a different perspective on the latter's extremity in terms of a maximal set C of priors.

The Ellsberg paradox is easily accommodated by the multiple priors model, once a suitable set C of priors is chosen. We have

$$W(\boldsymbol{a}_{\mathrm{II}}) = \min_{\mu \in \mathsf{C}} \sum_{\theta \in \Theta} m_\theta(BB \cup WB)\mu(\theta) = \frac{1}{100} \min_{\mu \in \mathsf{C}} \sum_{\theta=0}^{100} \theta\mu(\theta)$$

and

$$W(\boldsymbol{b}_{\mathrm{II}}) = \min_{\mu \in \mathsf{C}} \sum_{\theta \in \Theta} m_\theta(BW \cup WW)\mu(\theta) = 1 - \frac{1}{100} \max_{\mu \in \mathsf{C}} \sum_{\theta=0}^{100} \theta\mu(\theta).$$

Suppose there are priors $\mu, \mu' \in \mathsf{C}$ with expected number of black balls such that $\sum_{\theta=0}^{100} \theta\mu(\theta) < 50 < \sum_{\theta=0}^{100} \theta\mu'(\theta)$. Then both $W(\boldsymbol{a}_{\mathrm{II}}) < 1/2$ and $W(\boldsymbol{b}_{\mathrm{II}}) < 1/2$,[102] so the DM prefers to bet on the physical urn. For instance, suppose C consists of all

100. In the case of a DM who seeks uncertainty, $\succsim$ replaces $\precsim$ in equation (55) and the representation would feature a max instead of a min.

101. See Gajdos et al. (2008) and Eichberger and Guerdjikova (2013) for models that explicitly relate C with some underlying objective information (the latter paper within the case-based framework of Gilboa and Schmeidler 2001). Jaffray (1989) and Dominiak and Lefort (2015) are early and recent works that discuss this kind of information.

102. Note that $W(\boldsymbol{a}_{\mathrm{II}}) = W(\boldsymbol{b}_{\mathrm{II}})$ requires $\min_{\mu \in \mathsf{C}} \sum_{\theta=0}^{100} \theta\mu(\theta) + \max_{\mu \in \mathsf{C}} \sum_{\theta=0}^{100} \theta\mu(\theta) = 100$.

possible priors; that is, let $\mathsf{C} = \Delta(M)$. By (58), we have

$$W(\boldsymbol{a}_{\mathrm{II}}) = \min_{\theta \in \Theta} m_\theta(BB \cup WB) = 0 = \min_{\theta \in \Theta} m_\theta(BW \cup WW) = W(\boldsymbol{b}_{\mathrm{II}})$$

and so in this case we actually obtain the classical Ellsberg pattern (24).

5.2. Two-Stage Multiplier and Variational Preferences

We observed in Section 4.3 that, when $\phi(x) = -e^{-\lambda x}$, we have

$$\mathsf{c}_\lambda(\boldsymbol{a}, \mu) = -\frac{1}{\lambda} \log \int_M e^{-\lambda U(\boldsymbol{a},m)} d\mu(m)$$

which is a form of the multiplier preferences of Hansen and Sargent (2001, 2008).[103] As is well known (e.g., Cerreia-Vioglio et al. 2011),

$$-\frac{1}{\lambda} \log \int_M e^{-\lambda U(\boldsymbol{a},m)} d\mu(m) = \inf_{\nu \ll \mu} \left(\sum_{m \in M} U(\boldsymbol{a}, m)\nu(m) + \alpha R(\nu || \mu) \right), \tag{59}$$

where R is the relative entropy. At the same time, we can write the maxmin criterion (56) as

$$\begin{aligned} W(\boldsymbol{a}) &= \min_{\mu \in \mathsf{C}} \int_M U(\boldsymbol{a}, m) d\mu(m) \\ &= \inf_{\mu \in \Delta(M)} \left(\int_M U(\boldsymbol{a}, m) d\mu(m) + \delta_{\mathsf{C}}(\mu) \right), \end{aligned}$$

where $\delta_{\mathsf{C}} : \Delta(M) \to [0, +\infty)$ is the indicator function of C in the sense of convex analysis, that is,

$$\delta_{\mathsf{C}}(\mu) = \begin{cases} 0 & \text{if } \mu \in \mathsf{C} \\ +\infty & \text{else.} \end{cases} \tag{60}$$

This observation suggests that we consider the general two-stage representation

$$W(\boldsymbol{a}) = \inf_{\mu \in \Delta(M)} \left(\int_M U(\boldsymbol{a}, m) d\mu(m) + c(\mu) \right),$$

where $c : \Delta(M) \to [0, +\infty)$ is any convex function with $\inf_{\mu \in M} c(\mu) = 0$. Maccheroni, Marinacci, and Rustichini (2006) call *variational* the preferences that admit such a representation. Variational preferences extend both the maxmin and the multiplier preferences; the latter turn out to be the intersection of smooth and variational preferences (see Cerreia-Vioglio et al. 2011).

Maccheroni, Marinacci, and Rustichini (2006) axiomatize variational preferences and show how the function c captures ambiguity attitudes in a simple way: a variational

103. Recall from Section 4.1 that $U(\boldsymbol{a}, m) = \int_S u(\boldsymbol{a}(s)) dm(s)$.

preference $\succsim_1$ is *more ambiguity averse* (in the sense of Ghirardato and Marinacci 2002) than a variational preference $\succsim_2$ if $c_1(\mu) \leq c_2(\mu)$for every $\mu \in \Delta(M)$.[104] Heuristically, the index c can be also viewed as a confidence weight on priors, where $c(\mu) > c(\mu')$ means that the DM gives prior μ' a higher weight than μ. In the dichotomic case (60) that characterizes the maxmin criterion only two weights can be given.

Finally, Strzalecki (2011) shows which axioms characterize multiplier preferences within the class of variational preferences, and Cerreia-Vioglio et al. (2013a) axiomatize the two-stage version of variational preferences that we present here.

5.3. No Trade

The no-trade result of Dow and Werlang (1992) holds for the multiple priors criterion. Assume risk neutrality. The no-trade condition (32) becomes

$$\min_{\mu \in \mathsf{C}} \ \mathrm{E}_{\bar{\mu}}(y) < p < \max_{\mu \in \mathsf{C}} \ \mathrm{E}_{\bar{\mu}}(y) \tag{61}$$

The larger the set C (and so the higher the perceived model uncertainty or the aversion to it), the larger the set of prices that result in no trade. It is important to observe that—since (as previously discussed) the multiple priors criterion (56) need not exhibit extreme attitudes—it follows that the scope of the multiple priors no-trade result is actually much broader than for its Waldean counterpart (28), which does exhibit extreme attitudes.

In the present non-Bayesian setup, a low trade result holds more generally for variational preferences. In reading this result, note that for the variational criterion we have $W(\varepsilon y) \leq -W(-\varepsilon y)$, with equality if y is crisp.

PROPOSITION 10. *Let $\varepsilon > 0$. There is at most ε-trade volume in asset y if*

$$\frac{W(\varepsilon y)}{\varepsilon} < p < -\frac{W(-\varepsilon y)}{\varepsilon}. \tag{62}$$

When W is the multiple priors criterion, for each $\varepsilon > 0$ the condition (62) reduces to equation (61). In fact,

$$\min_{\mu \in \mathsf{C}} \ \mathrm{E}_{\bar{\mu}}(y) = W(y) < p < -W(-y) = \max_{\mu \in \mathsf{C}} \ \mathrm{E}_{\bar{\mu}}(y).$$

Hence, there is no trade whatsoever and we are back to the Dow and Werlang (1992) result. The Dow and Werlang insight has been developed in many papers, starting with Epstein and Wang (1994) (see Guidolin and Rinaldi 2013).

104. We refer to Gilboa and Marinacci (2013) for a detailed discussion of comparative ambiguity attitudes.

Appendix: Proofs and Related Analysis

A.1. Proof of Proposition 2

Points (i)–(iii) are easily checked. As to point (iv), observe that if ϕ and u are concave, then $\mathtt{c}(\cdot,\mu)$, and so $c(\cdot,\mu)$, is quasi-concave. Under either translation invariance (which, by (ii), holds in the CARA case) or positive homogeneity (which, by (iii), holds in the CRRA case), then $c(\cdot,\mu)$ is concave.

A.2. Proof of Proposition 3

To prove Proposition 3, we need a preliminary lemma, due to Hardy, Littlewood and Polya 1934 (p. 94).

LEMMA A.1. *Let $x_1 \le x_2 \le x_3 \le x_4$ and let $\phi : [x_1, x_4] \longrightarrow \mathbb{R}$ be concave. Then,*

$$\frac{\phi(x_4)-\phi(x_3)}{x_4-x_3} \le \frac{\phi(x_2)-\phi(x_1)}{x_2-x_1}.$$

Now let $U(\boldsymbol{b},m'') \le U(\boldsymbol{a},m'') \le U(\boldsymbol{a},m') \le U(\boldsymbol{b},m')$. Since for all $m \ne m', m''$, $U(\boldsymbol{a},m) = U(\boldsymbol{b},m)$, we have that

$$U(\boldsymbol{a},m')\mu(m') + U(\boldsymbol{a},m'')\mu(m'') = U(\boldsymbol{b},m')\mu(m') + U(\boldsymbol{b},m'')\mu(m'').$$

We want to show that

$$\mu(m')(\phi(U(\boldsymbol{b},m')) - \phi(U(\boldsymbol{a},m'))) + \mu(m'')(\phi(U(\boldsymbol{a},m'')) - \phi(U(\boldsymbol{b},m''))) \ge 0.$$

Assume $\mu(m'), \mu(m'') \ge 0$. The preceding equality implies

$$\mu(m')(U(\boldsymbol{b},m') - U(\boldsymbol{a},m')) = \mu(m'')(U(\boldsymbol{a},m'') - U(\boldsymbol{b},m''))$$

so that we can rewrite the preceding inequality as

$$\frac{\mu(m'')\left(\phi\left(U\left(\boldsymbol{a},m''\right)\right) - \phi\left(U(\boldsymbol{b},m'')\right)\right)}{\mu(m'')\left(U\left(\boldsymbol{a},m''\right) - U(\boldsymbol{b},m'')\right)} \ge \frac{\mu(m')\left(\phi(U(\boldsymbol{b},m')) - \phi\left(U(\boldsymbol{a},m')\right)\right)}{\mu(m')(U(\boldsymbol{b},m') - U(\boldsymbol{a},m'))}$$

or equivalently

$$\frac{\phi(U(\boldsymbol{b},m')) - \phi(U(\boldsymbol{a},m'))}{U(\boldsymbol{b},m') - U(\boldsymbol{a},m')} \le \frac{\phi(U(\boldsymbol{a},m'')) - \phi(U(\boldsymbol{b},m''))}{U(\boldsymbol{a},m'') - U(\boldsymbol{b},m'')}.$$

Now let us check the following cases.

1. If $U(\boldsymbol{b},m'') \le U(\boldsymbol{a},m'') \le U(\boldsymbol{a},m') = U(\boldsymbol{b},m')$, then the problem becomes one of verifying whether $\mu(m'')(\phi(U(\boldsymbol{a},m'')) - \phi(U(\boldsymbol{b},m''))) \ge 0$, which holds by the concavity of ϕ.

2. If $U(\boldsymbol{b}, m'') = U(\boldsymbol{a}, m'') \leq U(\boldsymbol{a}, m') \leq U(\boldsymbol{b}, m')$, then the problem is to verify
$$\mu(m')(\phi(U(\boldsymbol{b}, m')) - \phi(U(\boldsymbol{a}, m'))) \geq 0.$$
However, this is true because $U(\boldsymbol{b}, m'') = U(\boldsymbol{a}, m'')$ implies that $U(\boldsymbol{a}, m') = U(\boldsymbol{b}, m')$.
3. If $U(\boldsymbol{b}, m'') \leq U(\boldsymbol{a}, m'') = U(\boldsymbol{a}, m') \leq U(\boldsymbol{b}, m')$, then we have
$$\frac{\phi(U(\boldsymbol{b}, m')) - \phi(U(\boldsymbol{a}, m'))}{U(\boldsymbol{b}, m') - U(\boldsymbol{a}, m')} \leq \frac{\phi(U(\boldsymbol{a}, m')) - \phi\,(U(\boldsymbol{b}, m''))}{U(\boldsymbol{a}, m') - U(\boldsymbol{b}, m'')}$$
(see Hardy, Littlewood and Polya 1934, p. 93).
4. If $\mu(m') = 0$, then again we must verify $\mu(m'')(\phi(U(\boldsymbol{a}, m'')) - \phi(U(\boldsymbol{b}, m''))) \geq 0$, which holds by the concavity of ϕ .
5. If $\mu(m'') = 0$, then again we must verify $\mu(m')(\phi(U(\boldsymbol{b}, m')) - \phi(U(\boldsymbol{a}, m'))) \geq 0$. Yet this holds because $\mu(m'') = 0$ implies that $U(\boldsymbol{a}, m') = U(\boldsymbol{b}, m')$.
6. If $\mu(m') = \mu(m'') = 0$, then the inequality trivially holds as an equality to zero.

A.3. Proof of Proposition 4

If $x > 0$, then
$$\min_{m \in \operatorname{supp} \mu} \mathrm{E}_m(y) < p \iff V(\boldsymbol{x}) = \min_{m \in \operatorname{supp} \mu} \mathrm{E}_m(yx - px) < 0.$$
If $x < 0$, then
$$\begin{aligned} V(\boldsymbol{x}) &= \min_{m \in \operatorname{supp} \mu} \mathrm{E}_m(yx - xp) = - \max_{m \in \operatorname{supp} \mu} \mathrm{E}_m((-x)(y - p)) \\ &= x \max_{m \in \operatorname{supp} \mu} \mathrm{E}_m(y - p) \qquad \text{(A.1)} \end{aligned}$$
and so $\max_{m \in \operatorname{supp} \mu} \mathrm{E}_m(y) > p$ if and only if $V(\boldsymbol{x}) < 0$. We conclude that equation (32) holds if and only if $V(\boldsymbol{x}) < 0$ if $x \neq 0$.

A.4. Proof of Lemma 1

Since $\mathsf{c}_\lambda(\cdot, \mu)$ is concave, it follows that $\mathsf{c}_\lambda(\boldsymbol{x}, \mu) \leq -\mathsf{c}_\lambda(-\boldsymbol{x}, \mu) = \mathsf{c}_{-\lambda}(\boldsymbol{x}, \mu)$ for all $x \in \mathbb{R}$. If $\boldsymbol{x}$ is crisp, then $\mathsf{c}_\lambda(\boldsymbol{x}, \mu) = \mathrm{E}_m \boldsymbol{x} = - \mathsf{c}\,_\lambda(-\boldsymbol{x}, \mu)$.

A.5. Proof of Proposition 5

By Proposition 2(iv), c_λ is subhomogeneous. Fix any $x \in \mathbb{R}$. If $x \geq \varepsilon > 0$, there is a $k \geq 1$ such that $x = k\varepsilon$; hence
$$c_\lambda(\boldsymbol{x}, \mu) = c_\lambda(x(y - p), \mu) = c_\lambda(k\varepsilon(y - p), \mu) \leq k c_\lambda(\varepsilon(y - p), \mu).$$

Because

$$c_\lambda(\varepsilon(y-p),\mu) = c_\lambda(\varepsilon y,\mu) - \varepsilon p < 0 \Longleftrightarrow c_{\varepsilon\lambda}(y,\mu) = \frac{c_\lambda(\varepsilon y,\mu)}{\varepsilon} < p$$

we have that $c_{\varepsilon\lambda}(y,\mu) < p$ implies $c_\lambda(\boldsymbol{x},\mu) < 0$ when $x \geq \varepsilon$.

If $x \leq -\varepsilon < 0$, then there is a $k \geq 1$ such that $x = k(-\varepsilon)$, so

$$\begin{aligned} c_\lambda(\boldsymbol{x},\mu) &= c_\lambda(x(y-p),\mu) = c_\lambda(k(-\varepsilon)(y-p),\mu) \\ &= c_\lambda(k\varepsilon(p-y),\mu) \leq kc_\lambda(\varepsilon(p-y),\mu). \end{aligned}$$

We have

$$c_\lambda(\varepsilon(p-y),\mu) = \varepsilon p + c_\lambda(\varepsilon(-y),\mu) < 0 \Longleftrightarrow p < -\frac{c_\lambda(\varepsilon(-y),\mu)}{\varepsilon} = c_{-\varepsilon\lambda}(y,\mu).$$

Hence, $p < c_{-\varepsilon\lambda}(y,\mu)$ implies $c_\lambda(\boldsymbol{x},\mu) < 0$ when $x \leq -\varepsilon$. We conclude that (34) implies $c_\lambda(\boldsymbol{x},\mu) < 0$ for all $x \notin (-\varepsilon,\varepsilon)$.

A.6. Proof of Proposition 6

LEMMA A.2. *If* $\boldsymbol{a} \sim_{E_m} \boldsymbol{b}_m$ *for all* $m \in M$, *then* $\boldsymbol{a} \sim \sum_{m\in M} \boldsymbol{b}_m 1_{E_m}$.

Proof. Note that, $\boldsymbol{a} \sim_{E_m} \boldsymbol{b}_m$ if and only if, for every $\boldsymbol{c} \in A$,

$$\begin{bmatrix} \boldsymbol{c}(s) & \text{if } s \in E_1 \\ \vdots & \vdots \\ \boldsymbol{c}(s) & \text{if } s \in E_{m-1} \\ \boldsymbol{a}(s) & \text{if } s \in E_m \\ \boldsymbol{c}(s) & \text{if } s \in E_{m+1} \\ \vdots & \vdots \\ \boldsymbol{c}(s) & \text{if } s \in E_n \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{c}(s) & \text{if } s \in E_1 \\ \vdots & \vdots \\ \boldsymbol{c}(s) & \text{if } s \in E_{m-1} \\ \boldsymbol{b}_m(s) & \text{if } s \in E_m \\ \boldsymbol{c}(s) & \text{if } s \in E_{m+1} \\ \vdots & \vdots \\ \boldsymbol{c}(s) & \text{if } s \in E_n \end{bmatrix}$$

where $M = \{m_i\}_{i=1}^n$. If we omit (for brevity) the second column in the representation of acts above, repeated application of Assumption 3 delivers

$$\begin{bmatrix} \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \\ \vdots \\ \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{b}_1(s) \\ \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \\ \vdots \\ \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{b}_1(s) \\ \boldsymbol{b}_2(s) \\ \boldsymbol{a}(s) \\ \vdots \\ \boldsymbol{a}(s) \\ \boldsymbol{a}(s) \end{bmatrix} \sim \cdots \sim \begin{bmatrix} \boldsymbol{b}_1(s) \\ \boldsymbol{b}_2(s) \\ \boldsymbol{b}_3(s) \\ \vdots \\ \boldsymbol{b}_{M-1}(s) \\ \boldsymbol{a}(s) \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{b}_1(s) \\ \boldsymbol{b}_2(s) \\ \boldsymbol{b}_3(s) \\ \vdots \\ \boldsymbol{b}_{M-1}(s) \\ \boldsymbol{b}_M(s) \end{bmatrix}$$

The transitivity of $\sim$ now implies the result. $\square$

Now we proceed with the proof of Proposition 6. By Assumption 4, $\boldsymbol{a} \sim_{E_m} u^{-1}(\int_S u(\boldsymbol{a}(s))dm)$ for all $\boldsymbol{a} \in \boldsymbol{A}$ and $m \in M$. By the previous lemma, $\boldsymbol{a} \sim \sum_{m\in M}[u^{-1}(\int_S u(\boldsymbol{a}(s))dm(s))]1_{E_m}$. By Assumption 2, since $\sum_{m\in M}[u^{-1}(\int_S u(\boldsymbol{a})dm)]1_{E_m}$ is $\mathcal{E}$ measurable, we have

$$\boldsymbol{a} \sim \sum_{m\in M}\left[u^{-1}\left(\int_S u(\boldsymbol{a})dm\right)\right]1_{E_m} \sim v^{-1}\left(\sum_{m\in M} v\left[u^{-1}\left(\int_S u(\boldsymbol{a})dm\right)\right]P(E_m)\right).$$

Setting $\mu(m) = P(E_m)$, we conclude that

$$\boldsymbol{a} \sim v^{-1}\left(\int_M v\left[u^{-1}\left(\int_S u(\boldsymbol{a}(s))dm\right)\right]\mu(m)\right)$$

as desired.

A.7. *Proof of Proposition 7*

By standard arguments, there exists a unique solution $\hat{a} \in A$ for the problem $\max_{a\in A}\int_M \phi(R(a,m))d\mu(m)$. It satisfies the variational inequality (see, e.g., Kinderlehrer and Stampacchia 1980, Chap. 1)[105]

$$\nabla V(\hat{a})\cdot(a-\hat{a}) = \sum_{i=1}^{n}(a_i-\hat{a}_i)\int_M \phi'(R(\hat{a},m))\frac{\partial R}{\partial a_i}(\hat{a},m)d\mu(m) \le 0 \quad \forall a \in A.$$

Thus, since $\phi' > 0$, we have

$$\sum_{i=1}^{n}(a_i-\hat{a}_i)\int_M \frac{\phi'(R(\hat{a},m))}{\int_M \phi'(R(\hat{a},m))d\mu(m)}\frac{\partial R}{\partial a_i}(\hat{a},m)d\mu(m) \le 0 \qquad \forall a \in A.$$

Consider the function ζ given by (43). Since $\phi' > 0$, it holds $\zeta(m) > 0$ for each $m \in M$; moreover, $\sum_{m\in M}\zeta(m)\mu(m) = 1$. Define $\hat{\mu} \in \Delta(M)$ by $\hat{\mu}(m) = \zeta(m)\mu(m)$. Since $\zeta > 0$, $\hat{\mu}$ and μ are equivalent. We have $\nabla V(\hat{a})\cdot(a-\hat{a}) \le 0$ if and only if

$$\sum_{i=1}^{n}(a_i-\hat{a}_i)\int_M \frac{\partial R}{\partial a_i}(\hat{a},m)d\hat{\mu}(m) \le 0 \qquad \forall a \in A. \tag{A.2}$$

Yet because R is strictly concave, there is a unique solution $\bar{a} \in A$ for the problem

$$\max_{a\in A}\int_M R(a,m)d\hat{\mu}(m) \tag{A.3}$$

105. If $\hat{a}$ is an interior point, condition $\nabla V(\hat{a})\cdot(a-\hat{a}) \le 0$ reduces to $\nabla V(\hat{a}) = 0$.

and it is the unique action (Kinderlehrer and Stampacchia 1980, Chap. 1) that satisfies the variational inequality

$$\sum_{i=1}^{n}(a_i - \bar{a}_i)\int_M \frac{\partial R}{\partial a_i}(\bar{a}, m)d\hat{\mu}(m) \leq 0.$$

In view of (A.2), we conclude that $\hat{a} = \bar{a}$ and so $\hat{a}$ also solves problem (A.3).

A.8. Proof of Lemma 2

If $\hat{\mu} = \mu$, we have $\alpha = \min_{m\in\operatorname{supp}\mu} \zeta(m) = 1$. Suppose $\hat{\mu} \neq \mu$ and let $0 < \bar{\alpha} \leq \min_{m\in\operatorname{supp}\mu} \zeta(m)$. We have $\mu(m)\bar{\alpha} \leq \hat{\mu}(m)$ and so $\bar{\alpha} < 1$ because $\hat{\mu} \neq \mu$. Setting $\nu = (1/(1-\bar{\alpha}))\hat{\mu} - (\bar{\alpha}/(1-\bar{\alpha}))\mu \in \Delta(M)$, we can write $\hat{\mu} = \bar{\alpha}\mu + (1-\bar{\alpha})\nu$. As for the converse, let $\alpha \in (0,1]$ and $\nu \in \Delta(M)$ be such that $\hat{\mu} = \alpha\mu + (1-\alpha)\nu$. Then $\zeta(m)\mu(m) = \hat{\mu}(m) \geq \alpha\mu(m)$ for all $m \in M$, and so $\alpha \leq \min_{m\in\operatorname{supp}\mu} \zeta(m)$.

A.9. Proof of Proposition 8

From the constraints $c_0 = I_0 - \mathrm{E}_P(\rho w)$ and $c_{1i} = I_{1i} + w_i$, it follows that

$$c_0 \geq 0 \Longleftrightarrow \mathrm{E}_P(\rho w) \leq I_0 \quad \text{and} \quad c_{i1} \geq 0 \Longleftrightarrow w_i \geq -I_{1i}\,\forall i,$$

that is, $p_w \leq I_0$ and $w \geq -I$. Setting $A = \{w \in \mathbb{R}^k : \mathrm{E}_P(\rho w) \leq I_0 \text{ and } w \geq -I_1\}$, we can define the function $g : A \subseteq \mathbb{R}^k \to \mathbb{R}$ as

$$g(w) = \sum_m \phi\left(\sum_{i=1}^{k} u\left(I_0 - \mathrm{E}_{m^*}(\rho w), I_{1i} + w_i\right) m_i\right)\mu(m).$$

We can thus solve the optimization problem $\max_{w\in A} g(w)$. Since g is strictly concave, there is at most a solution $\hat{w}$. Suppose it is an interior point of A (so that the corresponding $\hat{c}$ is interior). The function g is differentiable at $\hat{w}$. Denote by $g'(\hat{w};\cdot) : W \to \mathbb{R}$ the directional derivative $g'(\hat{w}; w) = \lim_{t\downarrow 0}(g(\hat{w} + tw) - g(\hat{w}))/t$. Since $\hat{w}$ is interior, $g'(\hat{w}; w) = 0$ for every $w \in W$. By the chain rule,

$$\begin{aligned} 0 = g'(\hat{w}; w) = &-\mathrm{E}_\mu(\phi'(\mathrm{E}_m(u(\hat{c}))))\frac{\partial u}{\partial c_0}(\hat{c})\mathrm{E}_{m^*}(\rho w) \\ &+ \mathrm{E}_\mu\left(\phi'(\mathrm{E}_m(u(\hat{c})))\mathrm{E}_m\left(\frac{\partial u}{\partial c_1}(\hat{c})w\right)\right). \end{aligned}$$

By equation (46), $p_w = \mathrm{E}_{m^*}(\rho w)$, and so equation (50) follows.

A.10. Proof of Proposition 9

Define $\nu \in \Delta(S)$ by

$$\nu_i = \mathrm{E}_{\hat{\mu}}\left(\frac{\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}{\mathrm{E}_{\hat{\mu}}\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}\hat{m}_i\right) \qquad \forall i = 1,\ldots,k.$$

We have

$$\begin{aligned}
\nu_i &= \frac{\mathrm{E}_{\hat{\mu}}\left(\hat{m}_i \mathrm{E}_m(\partial u/\partial c_1)(\hat{c})\right)}{\mathrm{E}_{\hat{\mu}}\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})} = \sum_m \frac{\mathrm{E}_m(\partial u/\partial c_1)(c)}{\mathrm{E}_\mu\left(\zeta(m)\mathrm{E}_m(\partial u/\partial c_1)(c)\right)}\hat{m}_i\hat{\mu}(m) \\
&= \sum_m \frac{\mathrm{E}_m(\partial u/\partial c_1)(c)\mathrm{E}_\mu(\phi'(\mathrm{E}_m u(\hat{c})))}{\mathrm{E}_\mu\left(\phi'(\mathrm{E}_m u(\hat{c}))\mathrm{E}_m(\partial u/\partial c_1)(c)\right)}\hat{m}_i\hat{\mu}(m) \\
&= \sum_m \frac{\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})\mathrm{E}_\mu(\phi'(\mathrm{E}_m u(\hat{c})))}{\mathrm{E}_\mu\left(\phi'(\mathrm{E}_m u(\hat{c}))\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})\right)}\frac{(\partial u/\partial c_{1i})(\hat{c})}{\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}m_i\hat{\mu}(m) \\
&= \sum_m \frac{(\partial u/\partial c_{1i})(\hat{c})\mathrm{E}_\mu(\phi'(\mathrm{E}_m u(\hat{c})))}{\mathrm{E}_\mu\left(\phi'(\mathrm{E}_m u(\hat{c}))\mathrm{E}_m(\partial u/\partial c_1)(c)\right)}m_i\hat{\mu}(m) \\
&= \sum_m \frac{\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}\mathrm{E}_\mu(\phi'(\mathrm{E}_m u(\hat{c})))}{\mathrm{E}_\mu\left(\phi'(\mathrm{E}_m u(\hat{c}))\mathrm{E}_m\frac{(\partial u/\partial c_1)(c)}{(\partial u/\partial c_0)(\hat{c})}\right)}m_i\hat{\mu}(m) \\
&= \sum_m \frac{\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}}{\mathrm{E}_\mu\left(\zeta(m)\mathrm{E}_m\frac{(\partial u/\partial c_1(c))}{(\partial u/\partial c_0)(\hat{c})}\right)}m_i\hat{\mu}(m) \\
&= \sum_m \frac{\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}}{\mathrm{E}_{\hat{\mu}}\mathrm{E}_m\frac{(\partial u/\partial c_1)(c)}{(\partial u/\partial c_0)(\hat{c})}}m_i\hat{\mu}(m) = \frac{1}{\mathrm{E}_{m^*}\rho}\sum_m \frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u)(\partial c_0)(\hat{c})}m_i\hat{\mu}(m) \\
&= \frac{1}{\mathrm{E}_{m^*}\rho}\mathrm{E}_{\hat{\mu}}\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}m_i.
\end{aligned}$$

Note that $\mathrm{E}_{\hat{\mu}}\mathrm{E}_m((\partial u/\partial c_1)(c)/(\partial u/\partial c_0)(\hat{c})) = \mathrm{E}_{m^*}\rho = 1/r_f$ because the risk-free claim **1** is replicable. Since

$$\nu_i = \sum_m \frac{\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}{\mathrm{E}_{\hat{\mu}}\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}\hat{m}_i\hat{\mu}(m) = \sum_m \hat{m}_i\tilde{\mu}(m)$$

we can write

$$\begin{aligned}
\mathrm{E}_{m^*}(\rho)\mathrm{E}_{\tilde{\mu}}(\mathrm{E}_{\hat{m}}w) &= \mathrm{E}_{m^*}(\rho)\mathrm{E}_{\nu}w = \mathrm{E}_{m^*}(\rho)\sum_{i=1}^{k}\mathrm{E}_{\hat{\mu}}\left(\frac{\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}{\mathrm{E}_{\hat{\mu}}\mathrm{E}_m(\partial u/\partial c_1)(\hat{c})}\hat{m}_i\right)w_i \\
&= \sum_{i=1}^{k}\mathrm{E}_{\hat{\mu}}\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}m_i w_i = \mathrm{E}_{\hat{\mu}}\sum_{i=1}^{k}\frac{(\partial u/\partial c_{1i})(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}m_i w_i \\
&= \mathrm{E}_{\hat{\mu}}\mathrm{E}_m\left(\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}w\right) = p_w
\end{aligned}$$

where the last equality follows from equation (50).

A.11. Proof of Lemma 3

The if part is obvious. As to the converse, suppose that fractional treatment is not optimal. We first remark that the concavity of V is easy to check. Hence $V(0) \neq V(1)$ because otherwise $V(a) \geq V(0) = V(1)$ for all $a \in (0,1)$. Suppose $V(0) < V(1)$, so that the optimal decision is $\hat{a} = 1$—that is, $V(1) > V(a)$ for every $a \in [0,1)$. By concavity, $V(a) \geq aV(1) + (1-a)V(0) \geq V(0)$ for every $a \in [0,1]$, and so $V(0) = \min_{a\in[0,1]} V(a)$. Set $\phi(a) = V(a) - V(0)/(V(1) - V(0))$. Then, $\phi(0) = 0$ and $\phi(1) = 1$. Since V, and so ϕ, is concave, given any $\bar{a} \in (0,1)$, we have $\phi(a) \leq \phi(\bar{a}) + \phi'_+(\bar{a})(a - \bar{a})$ for all $a \in [0,1]$. Hence, $1 = \phi(1) \leq \phi(\bar{a}) + \phi'(\bar{a})(1-\bar{a})$, which implies $V'_+(\bar{a}) = \phi'_+(\bar{a}) > 0$ since $\phi(\bar{a}) < \phi(1) = 1$. We conclude that V is strictly increasing. If $V(0) > V(1)$, a similar argument shows that V is strictly decreasing.

A.12. Proof of Proposition 10

For brevity we omit the proof. It is similar to the proof of Proposition 5 but with W in place of c_λ.

A.13. Miscellanea

Equation (30). This is a special case of equation (58) with $M = \{\delta_s : s \in E\}$.

Equation (39). Assume that $\boldsymbol{a}$ is finitely valued. For $\beta \neq 0$ small enough, we have (see, e.g., Billingsley 1995, p. 148)

$$\begin{aligned}
\mathsf{c}_\beta(\boldsymbol{a},\mu) &= -\frac{1}{\beta}\log \mathrm{E}_\mu e^{\beta \mathrm{E}_m(-\boldsymbol{a})} = -\frac{1}{\beta}\sum_{n=1}^{\infty}\frac{\beta^n}{n!}c_n(\mathrm{E}(-\boldsymbol{a}))\\
&= -\sum_{n=1}^{\infty}\frac{(-1)^n\beta^{n-1}}{n!}c_n(\mathrm{E}(\boldsymbol{a})) = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}\beta^{n-1}}{n!}c_n(\mathrm{E}(\boldsymbol{a}))\\
&= \mathrm{E}_{\bar{\mu}}\boldsymbol{a} - \frac{1}{2}\alpha\sigma_\mu^2(\mathrm{E}(\boldsymbol{a})) + \sum_{n=3}^{\infty}\frac{(-1)^{n+1}\beta^{n-1}}{n!}c_n(\mathrm{E}(\boldsymbol{a})).
\end{aligned}$$

In turn, this implies equation (40).

Equation (50). We have

$$\begin{aligned}
p_w &= \mathrm{E}_\mu\left(\frac{\phi'(\mathrm{E}_m(u(\hat{c})))}{\mathrm{E}_\mu(\phi'(\mathrm{E}_m(u(\hat{c}))))}\mathrm{E}_m\left(\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}w\right)\right)\\
&= \mathrm{E}_\mu\left(\mathrm{E}_m\left(\frac{\phi'(\mathrm{E}_m(u(\hat{c})))(\partial u/\partial c_1)(\hat{c})}{\mathrm{E}_\mu(\phi'(\mathrm{E}_m(u(\hat{c}))))(\partial u/\partial c_0)(\hat{c})}w\right)\right)\\
&= \mathrm{E}_\mu\left(\mathrm{E}_{m^*}\left(\frac{\phi'(\mathrm{E}_m(u(\hat{c})))(\partial u/\partial c_1)(\hat{c})}{\mathrm{E}_\mu(\phi'(\mathrm{E}_m(u(\hat{c}))))(\partial u/\partial c_0)(\hat{c})}\frac{m}{m^*}w\right)\right)\\
&= \mathrm{E}_{m^*}\left(\frac{\mathrm{E}_\mu\left(\phi'(\mathrm{E}_m(u(\hat{c})))(m/m^*)\right)}{\mathrm{E}_\mu(\phi'(\mathrm{E}_m(u(\hat{c}))))}\frac{(\partial u/\partial c_1)(\hat{c})}{(\partial u/\partial c_0)(\hat{c})}w\right).
\end{aligned}$$

Since the market is complete, by considering Arrow contingent claims we get equation (51).

Example 4. We have $\hat{a} \in (0,1)$ if and only if $\mathrm{E}_\mu\bar{c}_m(t_0)\bar{c}_m(t_1) < \min\{\mathrm{E}_\mu\bar{c}_m^2(t_0), \mathrm{E}_\mu\bar{c}_m^2(t_1)\}$. Since for all $a, b \in \mathbb{R}$ it holds that $2\min\{a,b\} = (a+b) - |a-b|$, this inequality holds if and only if

$$2\mathrm{E}_\mu\bar{c}_m(t_0)\bar{c}_m(t_1) < \mathrm{E}_\mu\bar{c}_m^2(t_0) + \mathrm{E}_\mu\bar{c}_m^2(t_1) - \left|\mathrm{E}_\mu\bar{c}_m^2(t_0) - \mathrm{E}_\mu\bar{c}_m^2(t_1)\right|.$$

That is, $|V(a) - V(b)| = |\mathrm{E}_\mu\bar{c}_m^2(t_0) - \mathrm{E}_\mu\bar{c}_m^2(t_1)| < \mathrm{E}_\mu d_m^2$ as desired.

Equation (57). Since $\{\delta_m : m \in M\} \subseteq \Delta(M)$, the inequality $\leq$ follows. However, $\int_S u(\boldsymbol{a}(s))dm \geq \min_{m\in M}\int_S u(\boldsymbol{a}(s))dm$ for all $m \in M$, and so $\int_M(\int_S u(\boldsymbol{a}(s))dm)d\mu(m) \geq \min_{m\in M}\int_S u(\boldsymbol{a}(s))dm$ for all $\mu \in \Delta(M)$. Hence, also the inequality $\geq$ holds.

References

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker (2011). "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation." *American Economic Review*, 101(2), 695–723.

Akaike, Hirotugu (1985). "Prediction and Entropy." In *A Celebration of Statistics*, edited by Anthony C. Atkinson and Stephen E. Fienberg. Springer.

Amarante, Massimiliano (2009). "Foundations of Neo-Bayesian Statistics." *Journal of Economic Theory*, 144, 2146–2173.

Ang, Alfredo H. S. and Wilson H. Tang (2007). *Probability Concepts in Engineering,* 2nd ed. Wiley.

Anscombe, Francis J. and Robert J. Aumann (1963). "A Definition of Subjective Probability." *Annals of Mathematics and Statistics*, 34, 199–205.

Apostolakis, George (1990). "The Concept of Probability in Safety Assessments of Technological Systems." *Science*, 250, 1359–1364.

Arrow, Kenneth J. (1951). "Alternative Approaches to the Theory of Choice in Risk-Taking Situations." *Econometrica*, 19, 404–437.

Arrow, Kenneth J. and Leonid Hurwicz (1972). "An Optimality Criterion for Decision-Making under Ignorance." In *Uncertainty and Expectations in Economics*, edited by C. F. Carter and J. L. Ford. Basil Blackwell.

Aumann, Robert J. (1987). "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica*, 55, 1–18.

Backus, David, Axelle Ferriere, and Stanley Zin (2015). "Risk and Ambiguity in Models of Business Cycles." *Journal of Monetary Economics*, 69, 42–63.

Baker, Jack W. (2008). "An Introduction to Probabilistic Seismic Hazard Analysis (PSHA)." Working paper, Stanford University.

Barberis, Nicholas (2000). "Investing for the Long Run When Returns are Predictable." *Journal of Finance*, 55, 225–264.

Barron, Andrew, Jorma Rissanen, and Bin Yu (1998). "The Minimum Description Length Principle in Coding and Modeling." *IEEE Transactions on Information Theory*, 44, 2743–2760.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci (2015a). "Selfconfirming Equilibrium and Model Uncertainty." *American Economic Review*, 105(2), 646–677.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci, and Thomas Sargent (2015b). "Monetary Policies in Self-confirming Equilibria with Uncertain Models." Working paper, Bocconi University.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein (1979). *Estimation Risk and Optimal Portfolio Choice*. North-Holland.

Berger, James O. (1990). "Robust Bayesian Analysis: Sensitivity to the Prior." *Journal of Statistical Planning and Inference*, 25, 303–328.

Berger, James O. (1993). *Statistical Decision Theory and Bayesian Analysis,* 2nd ed. Springer.

Berger, James O. (2006). "The Case for Objective Bayesian Analysis." *Bayesian Analysis*, 1, 385–402.

Berger, Loïc, Johannes Emmerling, and Massimo Tavoni (2014). "Managing Catastrophic Climate Risks under Model Uncertainty Aversion." Working paper, Fondazione Eni Enrico Mattei (FEEM) and Euro-Mediterranean Center on Climate Change.

Bernardo, José M. and Adrian Smith (1994). *Bayesian Theory*. Wiley.

Bernoulli, Daniel (1954). "Specimen theoriae novae de mensura sortis." *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 1738 (translated in *Econometrica*, 22, 23–36).

Billingsley, Patrick (1965). *Ergodic Theory and Information*. Wiley.

Billingsley, Patrick (1995). *Probability and Measure*, 3rd ed. Wiley.

Boole, George (1854). *An Investigation in the Laws of Thought*. Walton and Maberley, London.

Borel, Emile (1909). *Elements de la théorie des probabilités*. Hermann, Paris.

Borgonovo, Emanuele, Veronica Cappelli, Fabio Maccheroni, and Massimo Marinacci (2015). "Risk Analysis and Decision Theory." Working paper 556, IGIER, Bocconi University.

Breiman, Leo (2001). "Statistical Modeling: The Two Cultures." *Statistical Science*, 16, 199–231.

Brock, William A., Steven N. Durlauf, and Kenneth D. West (2003). "Policy Evaluation in Uncertain Economic Environments." *Brookings Papers on Economic Activity*, 1, 235–322.

Carnap, Rudolf (1945). "The Two Concepts of Probability." *Philosophy and Phenomenological Research*, 5, 513–532.

Carnap, Rudolf (1950). *Logical Foundations of Probability*. University of Chicago Press.

Cerreia-Vioglio, Simone, Fabio Maccheroni, and Massimo Marinacci (2015). Put-Call Parity and Market Frictions." *Journal of Economic Theory*, 157, 730–762.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio (2011). "Uncertainty Averse Preferences." *Journal of Economic Theory*, 146, 1275–1330.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio 2013a. "Ambiguity and Robust Statistics." *Journal of Economic Theory*, 148, 974–1049.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio 2013b. "Classical Subjective Expected Utility." *Proceedings of the National Academy of Sciences*, 110, 6754–6759.

Chateauneuf, Alain and José H. Faro (2009). "Ambiguity through Confidence Functions." *Journal of Mathematical Economics*, 45, 535–558.

Chernoff, Herman (1954). "Rational Selection of Decision Functions." *Econometrica*, 22, 422–443.

Chew, Soo Hong and Jacob Sagi (2008). "Small Worlds: Modeling Attitudes toward Sources of Uncertainty." *Journal of Economic Theory*, 139, 1–24.

Cifarelli, Donato M. and Eugenio Regazzini (1996). "de Finetti's Contribution to Probability and Statistics." *Statistical Science*, 11, 253–282.

Cochrane, John H. (2005). *Asset Pricing*. Princeton University Press.

Cochrane, John H. and Lars P. Hansen (1992). "Asset Pricing Explorations for Macroeconomics." *NBER Macroeconomics Annual*, 7, 115–182.

Cohen, Michéle and Jean-Yves Jaffray (1980). "Rational Behavior under Complete Ignorance." *Econometrica*, 48, 1281–1299.

Collard, Fabrice, Sujoy Mukerji, Kevin Sheppard, and Jean-Marc Tallon (2012). "Ambiguity and the Historical Equity Premium." Working paper 500, Department of Economics, University of Oxford.

Cox, David R. (2006). *Principles of Statistical Inference*. Cambridge University Press.

Davis, Donald B. and M.-Elisabeth Paté-Cornell (1994). "A Challenge to the Compound Lottery Axiom: A Two-Stage Normative Structure and Comparison to Other Theories." *Theory and Decision*, 37, 267–309.

de Finetti, Bruno (1931). "Sul significato soggettivo della probabilità." *Fundamenta Mathematicae*, 17, 298–329.

de Finetti, Bruno (1937). "La prévision: ses lois logiques, ses sources subjectives." *Annales de l'Institut Henri Poincaré*, 7, 1–68.

de Finetti, Bruno (1971). "Probabilità di una teoria e probabilità dei fatti." In Studi in onore di G. Pompilj. Oderisi Editore, Gubbio.

de Finetti, Bruno (1977). Probabilities of Probabilities: A Real Problem or a Misunderstanding? In *New Developments in the Applications of Bayesian Methods*, edited by A. Aykac and C. Brumat. North-Holland.

Deming, William E. (1986). *Out of the Crisis: Quality, Productivity and Competitive Position*. Cambridge University Press.

Der Kiureghian, Armen (2008). "Analysis of Structural Reliability under Parameter Uncertainty." *Probabilistic Engineering Mechanics*, 23, 351–358.

Der Kiureghian, Armen and Ove Ditlevsen (2009). "Aleatory or Epistemic? Does It Matter?" *Structural Safety*, 31, 105–112.

Diaconis, Persi and David Freedman (1987). "A Dozen de Finetti-Style Results in Search of a Theory." *Annales de l'IHP Probabilités et Statistiques*, 23, 397–423.

Ditlevsen, Ove and Henrik O. Madsen (2007). *Structural Reliability Methods*. Wiley.

Dominiak, Adam and Jean-Philippe Lefort (2015). "Ambiguity and Probabilistic Information." Working paper.
Dow, James and Sergio R. C. Werlang (1992). "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio." *Econometrica*, 60, 197–204.
Draper, David (1995). "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society, Series B*, 45–97.
Draper, David, James S. Hodges, Edward E. Leamer, Carl N. Morris, and Donald B. Rubin (1987). *A Research Agenda for Assessment and Propagation of Model Uncertainty*. RAND Corporation Note.
Eichberger, Jürgen and Ani Guerdjikova (2013). "Ambiguity, Data and Preferences for Information—A Case-Based Approach." *Journal of Economic Theory*, 148, 1433–1462.
Ekeland, Ivar (1993). *The Broken Dice*. University of Chicago Press.
Ellsberg, Daniel (1961). "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics*, 75, 643–669.
Epstein, Larry and Tan Wang (1994). "Intertemporal Asset Pricing under Knightian Uncertainty." *Econometrica*, 62, 283–322.
Ergin, Haluk and Faruk Gul (2009). "A Theory of Subjective Compound Lotteries." *Journal of Economic Theory*, 144, 899–929.
Field, Richard V. and Mircea Grigoriu (2007). "Model Selection in Applied Science and Engineering: A Decision-Theoretic Approach." *Journal of Engineering Mechanics*, 133, 780–791.
Fishburn, Peter C. (1981). "Subjective Expected Utility: A Review of Normative Theories." *Theory and Decision*, 13, 139–199.
Fisher, Ronald A. (1957). "The Underworld of Probability." *Sankhya*, 18, 201–210.
Friedman, Milton (1953). "The Methodology of Positive Economics." In *Essays in Positive Economics*. University of Chicago Press.
Friedman, Milton and Leonard J. Savage (1948). "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy*, 16, 279–304.
Gajdos, Thibaut, Takashi Hayashi, Jean-Marc Tallon, and Jean-Christophe Vergnaud (2008). "Attitude Toward Imprecise Information." *Journal of Economic Theory*, 140, 27–65.
Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004). "Differentiating Ambiguity and Ambiguity Attitude." *Journal of Economic Theory*, 118, 133–173.
Ghirardato, Paolo and Massimo Marinacci (2001). "Risk, Ambiguity, and the Separation of Utility and Beliefs." *Mathematics of Operations Research*, 4, 864–89.
Ghirardato, Paolo and Massimo Marinacci (2002). "Ambiguity Made Precise: A Comparative Foundation." *Journal of Economic Theory*, 102, 251–289.
Giere, Ronald N. (1975). "Epistemological Roots of Scientific Knowledge." In *Induction, Probability, and Confirmation*, edited by G,. Maxwell and R. M. Anderson, Jr. University of Minnesota Press.
Giere, Ronald N. (1985). "Philosophy of Science Naturalized." *Philosophy of Science*, 52, 331–356.
Gilboa, Itzhak (2009). *Theory of Decision under Uncertainty*. Cambridge University Press.
Gilboa, Itzhak, Offer Lieberman, and David Schmeidler (2010). "On the Definition of Objective Probabilities by Empirical Similarity." *Synthese*, 172, 79–95.
Gilboa, Itzhak and Massimo Marinacci (2013). "Ambiguity and the Bayesian Paradigm." In *Advances in Economics and Econometrics: Theory and Applications*, edited by D. Acemoglu, M. Arellano, and E. Dekel. Cambridge University Press.
Gilboa, Itzhak and David Schmeidler (1989). "Maxmin Expected Utility with a Non-unique Prior." *Journal of Mathematical Economics*, 18, 141–153.
Gilboa, Itzhak and David Schmeidler (2001). *A Theory of Case-Based Decisions*. Cambridge University Press.
Giraud, Raphaël (2014). "Second Order Beliefs Models of Choice under Imprecise Risk: Nonadditive Second Order Beliefs versus Nonlinear Second Order Utility." *Theoretical Economics*, 9, 779–816.
Gollier, Christian (2011). "Portfolio Choices and Asset Prices: The Comparative Statics of Ambiguity Aversion." *Review of Economic Studies*, 78, 1329–1344.

Gollier, Christian (2013). *Pricing the Planet's Future: The Economics of Discounting in an Uncertain World*. Princeton University Press.

Good, Irving J. (1959). "Kinds of Probability." *Science*, 129, 443–450.

Good, Irving J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press.

Grant, Simon and Ben Polak (2013). "Mean-Dispersion Preferences and Constant Absolute Uncertainty Aversion." *Journal of Economic Theory*, 148, 1361–1398.

Guerdjikova, Ani and Emanuela Sciubba (2015). "Survival with Ambiguity." *Journal of Economic Theory*, 155, 50–94.

Guidolin, Massimo and Francesca Rinaldi (2013). "Ambiguity in Asset Pricing and Portfolio Choice: A Review of the Literature." *Theory and Decision*, 74, 183–217.

Gul, Faruk and Wolfgang Pesendorfer, (2015) "Hurwicz Expected Utility and Subjective Sources." *Journal of Economic Theory*, 159, 465–488

Guo, Jia and Xiaoping Du (2007). "Sensitivity Analysis with Mixture of Epistemic and Aleatory Uncertainties." *Journal of the American Institute of Aeronautics and Astronautics*, 45, 2337–2349.

Haavelmo, Trygve (1944). "The Probability Approach in Econometrics." *Econometrica*, 12, Supplement.

Hacking, Ian (1975). *The Emergence of Probability*. Cambridge University Press.

Halevy, Alon, Peter Norvig, and Fernando Pereira (2009). "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems*, 24, 8–12.

Hanany, Eran and Peter Klibanoff (2009). "Updating Ambiguity Averse Preferences." *BE Journal of Theoretical Economics*, 9, November.

Hansen, Lars P. (2007). "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk." *American Economic Review*, 97(2), 1–30.

Hansen, Lars P. (2014). "Uncertainty Outside and Inside Economic Models." *Journal of Political Economy* 122, 945–987.

Hansen, Lars P. and Scott F. Richard (1987). "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models." *Econometrica*, 55, 587–613.

Hansen, Lars P. and Thomas J. Sargent (2001). "Robust Control and Model Uncertainty." *American Economic Review*, 91, 60–66.

Hansen, Lars P. and Thomas J. Sargent (2007). "Recursive Robust Estimation and Control without Commitment." *Journal of Economic Theory*, 136, 1–27.

Hansen, Lars P. and Thomas J. Sargent (2008). *Robustness*. Princeton University Press.

Hansen, Lars P. and Thomas J. Sargent (2010). "Fragile Beliefs and the Price of Uncertainty." *Quantitative Economics*, 1, 129–162.

Hansen, Lars P. and Thomas J. Sargent (2014). *Uncertainty Within Economic Models*. World Scientific.

Hansen, Mark H. and Bin Yu (2001). "Model Selection and the Principle of Minimum Description Length." *Journal of the American Statistical Association*, 96, 746–774.

Hardy, Godfrey H., John E. Littlewood, and George Pólya (1934). *Inequalities*. Cambridge University Press.

Hayashi, Takashi and Jianjun Miao (2011). "Intertemporal Substitution and Recursive Smooth Ambiguity Preferences." *Theoretical Economics*, 6, 423–472.

Heal, Geoffrey and Antony Millner (2014). "Uncertainty and Decision Making in Climate Change Economics." *Review of Environmental Economics and Policy*, 8, 120–137.

Heckman, James J. (2008). "Econometric Causality." *International Statistical Review*, 76, 1–27.

Hurwicz, Leonid (1951). "Some Specification Problems and Application to Econometric Models." *Econometrica*, 19, 343–344.

Jaffray, Jean-Yves. (1989). "Linear Utility Theory for Belief Functions." *Operations Research Letters*, 8, 107–112.

Jahan-Parvar, Mohammad R. and Hening Liu (2014). "Ambiguity Aversion and Asset Prices in Production Economies." *Review of Financial Studies*, 27, 3060–3097.

Jaynes, Edwin T. (1959). *Probability Theory in Science and Engineering*. Mobil Oil Co., Dallas, TX.

Jaynes, Edwin T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jewitt, Ian and Sujoy Mukerji (2011). "Ordering Ambiguous Acts." Working paper, Department of Economics, University of Oxford.

Ju, Nengjiu and Jianjun Miao (2012). "Ambiguity, Learning, and Asset Returns." *Econometrica*, 80, 559–591.

Kandel, Shmuel and Robert F. Stambaugh (1996). "On the Predictability of Stock Returns: An Asset-Allocation Perspective." *Journal of Finance*, 51, 385–424.

Kaplan, Stanley and B. John Garrick (1981). "On the Quantitative Definition of Risk." *Risk Analysis*, 1, 1–28.

Kaplan, Stan, Harold F. Perla, and Dennis C. Bley (1983). "A Methodology for Seismic Risk Analysis of Nuclear Power Plants." *Risk Analysis*, 3, 169–180.

Kass, Robert E. and Larry Wasserman (1996). "The Selection of Prior Distributions by Formal Rules." *Journal of the American Statistical Association*, 91, 1343–1370.

Kelly, Thomas (2008). "Evidence: Fundamental Concepts and the Phenomenal Conception." *Philosophy Compass*, 3, 933–955.

Kendall, Maurice G. (1946). *The Advanced Theory of Statistics*, 2nd ed. Griffin, London.

Keynes, John M. (1921). *A Treatise on Probability*. Macmillan.

Kinderlehrer, David and Guido Stampacchia (1980). *An Introduction to Variational Inequalities and Their Applications*. Academic.

Klibanoff, Peter (2013). "Thoughts on Policy Choice Under Ambiguity." *Journal of Institutional and Theoretical Economics*, 169, 134–138.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005). "A Smooth Model of Decision Making Under Ambiguity." *Econometrica*, 73, 1849–1892.

Knight, Frank H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin.

Konishi, Sadanori and Genshiro Kitagawa (2008). *Information Criteria and Statistical Modeling*. Springer.

Koopmans, Tjalling C. (1947). "Measurement Without Theory." *Review of Economic Statistics*, 29, 161–172.

Kording, Konrad (2007). "Decision Theory: What 'Should' the Nervous System Do?" *Science*, 318, 606–610.

Kuhn, Harold (1953). "Extensive Games and the Problem of Information." In *Contributions to the Theory of Games*, edited by H. Kuhn and A. Tucker. Princeton University Press.

Laffont, Jean-Jacques (1989). *The Economics of Uncertainty and Information*. MIT Press.

Lang, Matthias (2015). "First-Order and Second-Order Ambiguity Aversion." Working paper.

Laplace, Pierre-Simon (1812). *Théorie analytique des probabilités*. Courcier, Paris.

Le Cam, Lucien (1977). "A Note on Metastatistics or 'An Essay Toward Stating a Problem in the Doctrine of Chances'." *Synthese*, 36, 133–160.

LeRoy, Stephen F., and Larry D. Singell (1987). "Knight on Risk and Uncertainty." *Journal of Political Economy*, 95, 394–406.

Lindley, Dennis V. (2013). *Understanding Uncertainty*, 2nd ed. Wiley.

Lucas, Robert E., Jr (1977). "Understanding Business Cycles." In *Stabilization of the Domestic and International Economy*, edited by K. Brunner and A. H. Meltzer. North-Holland.

Maccheroni, Fabio, Massimo Marinacci, and Doriana Ruffino (2013). "Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis." *Econometrica*, 81, 1075–1113.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006). "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences." *Econometrica*, 74, 1447–1498.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2012). "Social Decision Theory: Choosing Within and Between Groups." *Review of Economic Studies*, 79, 1591–1636.

Machina, Mark J. (2004). "Almost-Objective Uncertainty." *Economic Theory*, 24, 1–54.

Manski, Charles F. (2009). "Diversified Treatment Under Ambiguity." *International Economic Review*, 50, 1013–1041.

Marinacci, Massimo (2002). "Learning from Ambiguous Urns." *Statistical Papers*, 43, 143–151.

Marschak, Jacob (1953). "Economic Measurements for Policy and Prediction." In *Studies in Econometric Method*, edited by W. Hood and T. J. Koopmans. Wiley.

Marschak, Jacob and Roy Radner (1972). *Economic Theory of Teams*. Yale University Press.

Marzocchi, Warner, Christopher Newhall, and Gordon Woo (2012). "The Scientific Management of Volcanic Crises." *Journal of Volcanology and Geothermal Research*, 247/248, 181–189.

Mauldin, R. Daniel, David Preiss, and Heinrich von Weizsacker (1983). "Orthogonal Transition Kernels." *Annals of Probability*, 11, 970–988.

Meinshausen, Malte, Nicolai Meinshausen, William Hare, Sarah C. B. Raper, Katja Frieler, Reto Knutti, David J. Frame, and Myles R. Allen (2009). "Greenhouse-Gas Emission Targets for Limiting Global Warming to 2°C." *Nature*, 458, 1158–1162.

Miao, Jianjun (2014). *Economic Dynamics in Discrete Time*. MIT Press.

Millner, Antony, Simon Dietz, and Geoffrey Heal (2013). "Scientific Ambiguity and Climate Policy." *Environmental and Resource Economics*, 55, 21–46.

Milnor, John (1954). "Games Against Nature." In *Decision Processes*, edited by R. M. Thrall, C. H. Coombs, and R. L. Davis. Wiley.

Morris, Stephen (1995). "The Common Prior Assumption in Economic Theory." *Economics and Philosophy*, 11, 227–253.

Muliere, Pietro and Giovanni Parmigiani (1993). "Utility and Means in the 1930s." *Statistical Science*, 8, 421–432.

Muth, John F. (1961). "Rational Expectations and the Theory of Price Movements." *Econometrica*, 29, 315–335.

National Research Council (2009). "Science and Decisions: Advancing Risk Assessment." National Academies Press, Washington, DC.

Nau, Robert F. (2006). "Uncertainty Aversion with Second-Order Utilities and Probabilities." *Management Science*, 52, 136–145 (see also *Proceedings of ISIPTA 2001*).

Neilson, William S. (2010). "A Simplified Axiomatic Approach to Ambiguity Aversion." *Journal of Risk and Uncertainty*, 41, 113–124.

Neyman, Jerzy (1957). "Inductive Behavior as a Basic Concept of Philosophy of Science." *Review of the International Statistical Institute*, 25, 7–22.

Ornstein, Donald S. and Benjamin Weiss (1990). "How Sampling Reveals a Process." *Annals of Probability*, 18, 905–930.

Pareto, Vilfredo (1900). "Sunto di alcuni capitoli di un nuovo trattato di economia pura." *Giornale degli Economisti*, 20, 216–235 (translated in *Giornale degli Economisti*, 67, 453–504, 2008).

Pastor, Luboš (2000). "Portfolio Selection and Asset Pricing Models." *Journal of Finance*, 55, 179–223.

Paté-Cornell, M.-Elisabeth, (1996). "Uncertainties in Risk Analysis: Six Levels of Treatment." *Reliability Engineering and System Safety*, 54, 95–111.

Pearl, Judea (1988). "*Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*." Morgan Kaufmann.

Picci, Giorgio (1977). "Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem." *SIAM Journal on Applied Mathematics*, 33, 383–398.

Popper, Karl R. (1959). "The Propensity Interpretation of Probability." *British Journal for the Philosophy of Science*, 10, 25–42.

Porter, Keith, Kim Shoaf, and Hope Seligson (2006). "Value of Injuries in the Northridge Earthquake." *Earthquake Spectra*, 22, 555–563.

Porteus, Evan L. (2002). "*Foundations of Stochastic Inventory Theory*." Stanford University Press.

Pratt, John W. and Robert Schlaifer (1984). "On the Nature and Discovery of Structure." *Journal of the American Statistical Association*, 79, 9–21.

Raiffa, Howard and Robert Schlaifer (1961). *Applied Statistical Decision Theory*. Harvard Graduate School of Business Administration.

Ramsey, Frank P. (1931). "Truth and Probability (1926)." In *The Foundations of Mathematics and Other Logical Essays*, edited by R. Braithwaite. Routledge and Kegan Paul.

Rissanen, Jorma (2007). *Information and Complexity in Statistical Modeling*. Springer.

Rogelj, Joeri, Malte Meinshausen, and Reto Knutti (2012). "Global Warming under Old and New Scenarios Using IPCC Climate Sensitivity Range Estimates." *Nature Climate Change*, 2, 248–253.

Ross, Stephen A. (2005). *Neoclassical Finance*. Princeton University Press.

Rothschild, Michael J. and Joseph Stiglitz (1970). "Increasing Risk I: A Definition." *Journal of Economic Theory*, 2, 225–243.
Samuelson, Paul A. (1938). "A Note on the Pure Theory of Consumer's Behaviour." *Economica*, 5, 61–71.
Sargent, Thomas J. (2008). "Evolution and Intelligent Design." *American Economic Review*, 98(1), 3–37.
Savage, Leonard J. (1951). "The Theory of Statistical Decision." *Journal of the American Statistical Association*, 46, 55–67.
Savage, Leonard J. (1954). *The Foundations of Statistics*. Wiley.
Schmeidler, David (1989). "Subjective Probability and Expected Utility Without Additivity." *Econometrica*, 57, 571–587.
Segal, Uzi (1987). "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach." *International Economic Review*, 28, 175–202.
Seo, Kyoungwon (2009). "Ambiguity and Second-Order Belief." *Econometrica*, 77, 1575–1605.
Shafer, Glenn (1978). "Non-additive Probabilities in the Work of Bernoulli and Lambert." *Archive for History of Exact Sciences*, 19, 309–370.
Singpurwalla, Nozer D. (2006). *Reliability and Risk: A Bayesian Perspective*. Wiley.
Siniscalchi, Marciano (2011). "Dynamic Choice Under Ambiguity." *Theoretical Economics*, 6, 379–421.
Skinner, Burrhus Frederic (1985). "Cognitive Science and Behaviourism." *British Journal of Psychology*, 76, 291–301.
Smith, Vernon L. (1969). "Measuring Nonmonetary Utilities in Uncertain Choices: The Ellsberg Urn." *Quarterly Journal of Economics*, 83, 324–329.
Solomon, Susan et al. (2007). *IPCC Climate Change 2007: The Physical Science Basis*. Cambridge University Press.
Stern, Nicholas (2007). *The Economics of Climate Change*. Cambridge University Press.
Stoye, Jörg (2011). "Statistical Decisions Under Ambiguity." *Theory and Decision*, 70, 129–148.
Strzalecki, Tomasz (2011). "Axiomatic Foundations of Multiplier Preferences." *Econometrica*, 79, 47–73.
Strzalecki, Tomasz (2013). "Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion." *Econometrica*, 81, 1039–1074.
Taboga, Marco (2005). "Portfolio Selection with Two-Stage Preferences." *Finance Research Letters*, 2, 152–164.
Teicher, Henry (1963). "Identifiability of Finite Mixtures." *Annals of Mathematical Statistics*, 34, 244–248.
Tenenbaum, Joshua B., Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman (2011). "How to Grow a Mind: Statistics, Structure, and Abstraction." *Science*, 331, 1279–1285.
Theil, Henri (1961). *Economic Forecasts and Policy*. North-Holland.
Tinbergen, Jan (1952). *On the Theory of Economic Policy*. North-Holland.
von Mises, Richard (1939). *Probability, Statistics, and Truth*. Macmillan.
von Neumann, John and Oskar Morgenstern (1947). *Theory of Games and Economic Behavior*, 2nd ed. Princeton University Press.
von Plato, Jan (1988). "Ergodic Theory and the Foundations of Probability." In *Causation, Chance, and Credence*, edited by B. Skyrms and W. L. Harper. Kluwer.
von Plato, Jan, (1989). "Probability in Dynamical Systems." In *Logic, Methodology and Philosophy of Science VIII*, edited by J,. E. Fenstad, I. T. Frolov, and R. Hilpinen. North-Holland.
von Plato, Jan (1994). *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*. Cambridge University Press.
Wakker, Peter P. (2010). *Prospect Theory*. Cambridge University Press.
Wald, Abraham (1950). *Statistical Decision Functions*. Wiley.
Walker, Warren E., Poul Harremoes, Jan Rotmans, Jan P. van der Sluijs, Marjolein Asselt, Peter Janssen, and Martin Krayer von Krauss (2003). "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." *Integrated Assessment*, 4, 5–17.

Weibull, Jörgen W. (1996). "The Mass Action Interpretation." Research Institute of Industrial Economics Working Paper No. 427. In *The Work of John Nash in Game Theory*, edited by Harold Kuhn et al. *Journal of Economic Theory*, 69, 153–185.

Winkler, Robert L. (1996). "Uncertainty in Probabilistic Risk Assessment." *Reliability Engineering and System Safety*, 54, 127–132.

Xu, Fei and Vashti Garcia (2008). "Intuitive Statistics by 8-Month-Old Infants." *Proceedings of the National Academy of Sciences*, 105, 5012–5015.

Xu, Fei and Tamar Kushnir (2013). "Infants are Rational Constructivist Learners." *Current Directions in Psychological Science*, 22, 28–32.

Xue, Jingyi (2012). "Three Representations of Preferences with Decreasing Absolute Uncertainty Aversion." Working paper, Rice University.

Zabell, Sandy L. (2011). "The Subjective and the Objective." *Handbook of the Philosophy of Science*, Vol. 7, edited by Prasanta S. Bandyopadyay and Malcolm R. Forster. Elsevier, pp. 1149–1174.