SYMMETRY OF EVIDENCE WITHOUT EVIDENCE OF SYMMETRY*

Larry G. Epstein Kyoungwon Seo

January 17, 2010

Abstract

The de Finetti Theorem is a cornerstone of the Bayesian approach. Bernardo [4, p. 5] writes that its "message is very clear: if a sequence of observations is judged to be exchangeable, then any subset of them must be regarded as a random sample from some model, and there exists a prior distribution on the parameter of such model, hence *requiring* a Bayesian approach." We argue that while exchangeability, interpreted as symmetry of evidence, is a weak assumption, when combined with subjective expected utility theory, it implies also complete confidence that experiments are identical. When evidence is sparse, and there is little evidence of symmetry, this implication of de Finetti's hypotheses is not intuitive. This motivates our adoption of multiple-priors utility as the benchmark model of preference. We provide two alternative generalizations of the de Finetti Theorem for this framework. A model of updating is also provided.

1. INTRODUCTION

1.1. Motivation and Objectives

An individual is considering bets on the outcomes of a sequence of coin tosses. It is the same coin being tossed repeatedly, but different tosses are performed by different people. The individual believes that outcomes depend on both the (unknown) physical make-up or bias of the coin and on the way in which the coin is tossed. Her understanding of tossing technique is poor. However, she has no reason to distinguish between the techniques of

*Boston University, lepstein@bu.edu and Department of Managerial Economics and Decision Sciences, Northwestern University, k-seo@kellogg.northwestern.edu. We are grateful to Asen Kochov, Fabio Maccheroni, Massimo Marinacci, Uzi Segal, participants at RUD 08 at Oxford, and especially to Martin Schneider for helpful comments and discussions, and to Yi-Chun Chen for providing us with his unpublished work which helped us to better understand "regularity" in the multiple-priors model. This research was supported by the National Science Foundation (awards SES-0611456, SES-0917740 and SES-0918248). First version: June 2008. This version incorporates (in Section 6) some results on updating initially reported in "Symmetry, ambiguity and frequencies," which no longer exists as a separate paper.

different people and she views technique as being idiosyncratic. Given this perception, how would she rank bets?

More generally, we are interested in modeling a decision-maker who is facing a sequence of experiments, and whose perception is that outcomes are influenced by two factors - one that is well understood and fixed across experiments (coin bias), and the other that is poorly understood and thought to be unrelated across experiments. This description would seem to apply to many choice settings, where the decision-maker has a theory or model of her environment, but where she is sophisticated enough to realize that it is "incomplete" - hence the second factor, which can be thought of as an "error term" for her model.

We limit ourselves to situations where, in addition, there is *symmetry of evidence* about the experiments - no information is given that would imply a distinction between them. However, if little information is provided about any of the experiments, in which case there is little *evidence of symmetry*, a sophisticated individual might very well admit the possibility that the experiments may differ in some way, and this may influence her ranking of bets. The distinction between the two forms of symmetry is due to Walley [37], who also argued that this distinction is behaviorally meaningful and that it cannot be accommodated within the Bayesian framework. Following the terminological distinction introduced in Epstein and Schneider [15], we also refer to experiments as being *indistinguishable* but *not* necessarily *identical*.

A prime motivating example is where the decision-maker is a statistician or empiricist, and an experiment is part of a statistical model of how data are generated. Invariably symmetry is assumed at some level - perhaps after correcting for perceived asymmetries, such as heteroscedasticity of errors in a regression model. Standard statistical methods presume that, after such corrections, the *identical* statistical model applies to all experiments or observations. This practice has been criticized as being particularly inappropriate in the context of the literature attempting to explain cross-country differences in growth rates, in which case an 'experiment" corresponds to a country. Brock and Durlauf [5, p. 231] argue that it is "a major source of skepticism about the empirical growth literature." They write further that "where the analyst can be specific about potential differences [between countries], she can presumably (test and) correct for them by existing statistical methods. However, the open-endedness of growth theories makes it impossible to account in this way for all possible differences." Since they also emphasize the importance of having sound decision-theoretic foundations for statistical methods, particularly for purposes of policy analysis, we interpret their paper as calling (first) for a model of decision-making that would permit the analyst to express a judgement of "similarity" or "indistinguishability," but also a concern that countries or experiments may differ, even if she cannot specify how. Such a model is our objective.

1.2. The De Finetti Bayesian Model

Some readers may be wondering why there is a need for a new model of choice - does not the exchangeable Bayesian model due to de Finetti adequately capture beliefs and, in conjunction with subjective expected utility, also choice, in the coin-tossing setting (and more generally)?

Recall de Finetti's model and celebrated theorem [11, 23]. There is a countable infinity of experiments, indexed by the set $\mathbb{N} = \{1, 2, ...\}$. Each experiment yields an outcome in the set S and thus $\Omega = S^\infty$ is the set of all possible sample paths (technical details are suppressed until later). A probability measure P on Ω is *exchangeable* if

$$P\left(A_1 \times A_2 \times .....\right) = P(A_{\pi^{-1}(1)} \times A_{\pi^{-1}(2)} \times .....),$$

for all finite permutations π of $\mathbb{N}$. De Finetti shows that exchangeability is equivalent to the following representation: There exists a (necessarily unique) probability measure μ on $\Delta\left(S\right)$ such that

$$P\left(\cdot\right) = \int_{\Delta(S)} \ell^\infty\left(\cdot\right) d\mu\left(\ell\right), \tag{1.1}$$

where, for any probability measure ℓ on S (written $\ell \in \Delta\left(S\right)$), ℓ^∞ denotes the corresponding i.i.d. product measure on Ω.[1]

Given a Bayesian prior, symmetry of evidence implies exchangeability and therefore de Finetti's representation, which admits the obvious interpretation: The individual is uncertain about which probability law ℓ describes any single experiment. However, conditional on any ℓ in the support of μ, it is the i.i.d. product ℓ^∞ that describes the implied probability law on Ω. This suggests that there is no room in the model to accommodate a concern with experiments not being identical. In Section 4, we confirm this suggestion at the behavioral level by *identifying behavior that is intuitive for an individual who is not completely confident that experiments are identical, but yet is ruled out by the Independence axiom* of subjective expected utility theory. Thus we propose a model that generalizes the exchangeable Bayesian model by suitably relaxing the Independence axiom.

Specifically, we adopt the framework of multiple-priors utility (Gilboa and Schmeidler [19]), and specialize it by adding axioms, forms of "exchangeability," for example, that capture alternative hypotheses about how the relationship between experiments is perceived. Two alternative generalizations of de Finetti's theorem are established. In the first (Theorem 3.2), the decomposition (1.1) of a Bayesian prior is generalized so that the individual's *set* of priors $\mathcal{P}$ has the form

$$\mathcal{P} = \left\{ \int \ell^\infty\left(\cdot\right) d\mu\left(\ell\right) : \mu \in \mathcal{M} \right\},$$

for some set of probability measures $\mathcal{M}$ over $\Delta\left(S\right)$; equivalently, every measure in $\mathcal{P}$ is exchangeable. An interpretation is that there is ex ante ambiguity about which likelihood function applies, but certainty that the same likelihood function applies to all experiments. Thus, just for the Bayesian case, experiments are perceived as identical.

The second generalization of de Finetti's Theorem, (see Theorem 5.2), relaxes the latter feature and accommodates the absence of (overwhelming) evidence of symmetry. (On the other hand, ex ante ambiguity is precluded, so that this result does not generalize the first one.) The corresponding representation for $\mathcal{P}$ is more complicated - it retains

[1] Though the de Finetti theorem can be viewed as a result in probability theory alone, it is typically understood in economics as describing the prior in the subjective expected utility model of choice. That is how we view it in this paper.

a counterpart of the single prior μ, but where μ is a probability measure over, roughly speaking, nonsingleton sets of likelihoods, which sets are the unknown parameters in the representation. The following informal description gives a sense of how "indistinguishable but not identical" is captured. Consider the introductory coin-tossing setting for concreteness, so that $S = \{H, T\}$. In the Bayesian model, each experiment is characterized by a single number in the unit interval - the probability of Heads. Here, instead an experiment is characterized by an interval of probabilities for Heads, which is nondegenerate because even given the physical bias of the coin, the influence of tossing technique is poorly understood. Experiments are indistinguishable, because each is described by the same interval. However, they are not identical, because any probability in the interval could apply to any experiment. The length of the interval parametrizes the importance of idiosyncratic poorly understood factors, and varies with preference, hence with the individual.

1.3. Updating

As indicated, one formal contribution of this paper is to generalize de Finetti's Theorem from probability measures to sets of priors. However, the importance of the de Finetti Theorem extends beyond the representation to the connection it affords between subjective beliefs and empirical frequencies, most notably through Bayesian updating of the prior μ. The combination of the de Finetti Theorem and Bayes' Rule gives the canonical model of learning or inference in economics and statistics. Under well-known conditions, it yields the important conclusion that priors will eventually be swamped by data and that individuals will learn the truth (see Savage [33, Ch. 3.6], for example). Our second major contribution is to show that (with some qualification) Bayesian updating extends to the case where experiments may not be identical, as formalized by our second model (Theorem 5.2).

It is well known that ambiguity poses difficulties for updating and that there is no consensus updating rule analogous to Bayes' Rule. However, our second model admits intuitive (and dynamically consistent) updating in a limited but still interesting class of environments, namely, where an individual first samples and observes the outcomes of some experiments, and then chooses how to bet on the outcomes of remaining experiments. The essential point is that each experiment serves *either* as a signal *or* is payoff relevant, but *not both*. For example, think of a statistical decision-maker who, after observing the results of some experiments, is concerned with predicting the results of others because he must take an action (estimation, or hypothesis testing perhaps) whose payoff depends on their outcomes. Policy evaluation in the context of cross-country growth is a concrete application, where choice between policies for a particular country is based on observations of how these policies fared in others. Our model prescribes a way to use the latter information that accommodates the policy-maker's concern that countries may differ in ways that are poorly understood and that are not taken into account in the model of growth.

Besides being well-founded axiomatically, the model of updating is also tractable. This aspect stems from the fact that given the model of Theorem 5.2, beliefs at every node are completely defined by a (unique) probability measure over the unknown parameters. Thus one need only describe how information is incorporated into an additive probability measure, rather than dealing with the thornier problem of updating a set of priors. As shown in Theorem 6.1, this can be done in a way that mirrors standard Bayesian updating. A consequence is that formal results from Bayesian learning theory can be translated into our model, though with suitable reinterpretation. As one example, we establish (Proposition 6.3) a counterpart of the Savage result that data eventually swamp the prior. In the coin-tossing example, the individual asymptotically converges to certainty about a particular bias, and hence about a specific probability interval, but since she may still be left with an interval, she may remain ambiguous about tossing technique and thus remain concerned that experiments differ. She learns all that she believes that she can, given her ex ante perception of the experiments, which, in turn, underlies her preferences. If the truth is that tossing technique is not important, and if that possibility is admitted in her prior view, then she will converge to the truth asymptotically.

1.4. Related Literature

Kreps [25, Ch. 11] refers to the de Finetti Theorem as "the fundamental theorem of (most) statistics," because of the justification it provides for the analyst to view samples as being independent and identically distributed with unknown distribution function - this is warranted if and only if samples are assessed ex ante as being exchangeable. As a result, and also because similarity judgements naturally play a central role in statistical analysis, the notion of exchangeability underlies much of common empirical practice.

Bayesians often refer to exchangeability as a weak assumption. Schervish [34, p. 8] writes: "The motivation for the definition of exchangeability is to express symmetry of beliefs ... in the weakest possible way. The definition ... does not require any judgement of independence or that any limit of relative frequencies will exist. It merely says that the labeling of random quantities is immaterial." We agree that "symmetry of beliefs", in the sense of "symmetry of evidence", is a weak assumption. Our objection is to the (implicit) companion hypothesis of SEU preferences. To improve upon exchangeability, Bayesians have proposed weaker notions that build in less symmetry, while maintaining SEU; see Schervish's Ch. 8, for example. *Such extensions within the Bayesian framework do not permit the separate modeling of a concern with evidence of symmetry in an environment where evidence is symmetric.*

Brock and Durlauf's [5] critique of the empirical growth literature is in part expressed as a critique of the assumption of (a conditional or partial form of) exchangeability. In our view, the culprit is not symmetry, but rather the implicit assumption of expected utility theory.

We have already acknowledged our debt to Walley [37] for the critique that motivates this paper and for the distinction that we have adopted as a title. His contribution to modeling the distinction is described briefly in Section 3.

Finally, Epstein and Schneider [15] model the distinction between symmetry of evidence and evidence of symmetry in the special case where experiments are viewed as being completely unrelated (in the context of the above example of repeated tosses of a single coin, they assume that the physical bias is known with certainty). In [16], those authors study the more general case dealt with here (unknown physical bias). A major difference from this paper is that they describe functional forms and provide informal justification, partly through applications, while here the focus is on axiomatic foundations.

2. PRELIMINARIES

2.1. The Bayesian Model

There exists a countable infinity of experiments - they are ordered and indexed by the set $\mathbb{N} = \{1, 2, ...\}$. Each experiment yields an outcome in the finite set S. The set of possible outcomes for the i^{th}experiment is sometimes denoted S_i, though $S_i = S$ for all i. The full state space is

$$\Omega = S^{\infty} = S_1 \times S_2 \times \; .... \; = S^{\infty}.$$

Denote by Σ the product σ-algebra on Ω. Probability measures on (Ω, Σ) are understood to be countably additive unless specified otherwise.

An *act* is a Σ-measurable function from Ω into $[0, 1]$. For example, when $S = \{H, T\}$, then the act f,

$$f(s_1, ..., s_i, ...) = \begin{cases} 1 & (s_1, s_2) = (H, T) \\ 0 & \text{otherwise,} \end{cases}$$

is the bet on Heads followed by Tails; below it will often be abbreviated by H_1T_2 (similar abbreviations are adopted for other acts in the coin-tossing context). Preference, denoted $\succeq$, is defined on the set $\mathcal{F}$ of all acts.

For any subset I of $\mathbb{N}$, Σ_I denotes the product σ-algebra on $\Pi_{i \in I} S_i$, also identified with a σ-algebra on Ω. Denote by $\mathcal{F}_I$ the set of all acts that are Σ_I-measurable. (When $I = \{i\}$, we write Σ_i and $f \in \mathcal{F}_i$.) Such acts will be said to depend only on experiments in I. Particularly important are acts that depend on finitely many experiments, that is, acts in

$$\mathcal{F}_{fin} = \cup_{I \text{ finite}} \mathcal{F}_I.$$

Refer to such acts as *finitely-based*.

Denote by Π the set of finite permutations of $\mathbb{N}$; all permutations appearing in the paper should be understood to be finite. For any π in Π and probability measure P on (S^{∞}, Σ), define πP to be the unique probability measure on S^{∞} satisfying (for all rectangles)

$$(\pi P)\,(A_1 \times A_2 \times .....) = P(A_{\pi^{-1}(1)} \times A_{\pi^{-1}(2)} \times .....).$$

Given an act f, define the permuted act πf by $(\pi f)(s_1, ..., s_t, ...) = f\left(s_{\pi(1)}, ..., s_{\pi(t)}, ...\right)$. Abbreviate $\int f dP$ by Pf, or $P(f)$. Then, for all P, f and π,

$$(\pi P) f = P(\pi f).$$

The probability measure P is *exchangeable* if $\pi P = P$ for all π. In behavioral terms, assuming subjective expected utility preference with prior P, exchangeability of P is equivalent to the universal indifference between an act and any permuted variant, that is,

$$f \sim \pi f \quad \text{for all acts } f \text{ and permutations } \pi.$$

For any probability measure ℓ on S (write $\ell \in \Delta(S)$), ℓ^{∞} denotes the corresponding i.i.d. product measure on (Ω, Σ).

Theorem 2.1 (de Finetti). *The probability measure P on (Ω, Σ) is exchangeable if and only if there exists a (necessarily unique) Borel probability measure μ on $\Delta(S)$ such that*

$$P(\cdot) = \int_{\Delta(S)} \ell^{\infty}(\cdot)\, d\mu(\ell).$$

A noteworthy and problematic feature of the framework, which we adopt also below, is that payoffs to acts depend on the outcomes of infinitely many experiments, which is problematic for a positive model. In particular, the domain of preference includes acts whose payoffs depend on the truth/falsity of tail events, which are not observed in finite time, and thus are, in fact, unobservable.[2] This concern was emphasized also by de Finetti; see [32] for extensive discussion of de Finetti's view, and also [12]. However, a decision-maker might be able to conceive of payoffs that depend on tail events (receive x^* if the limiting empirical frequency of Heads in an infinite sequence of tosses is greater than $\frac{1}{2}$ and x otherwise). Thus the de Finetti theorem and its generalizations below seem useful in a normative context.[3]

Another objection to the de Finetti-Savage model is the one raised by Walley and described in the introduction - that symmetry of evidence in their model implies also that experiments are necessarily viewed as being identical. Elaborating upon and accommodating this critique are the objectives of this paper, and is the reason that we move from subjective expected utility (SEU) to the multiple-priors model.

For any compact metric space X, $\Delta(X)$ denotes the set of countably additive Borel probability measures on X, endowed with the weak-convergence topology induced by continuous functions. $\mathcal{K}(X)$ denotes the space of compact subsets of X, endowed with the Hausdorff metric topology, which renders it compact metric. When X is a lts, $\mathcal{K}^c(X)$ denotes the subspace of compact and convex subsets of X.

2.2. Multiple-Priors Preference

By a multiple-priors preference (or utility), we shall mean a preference $\succeq$ on $\mathcal{F}$ that has a representation of the following form. There exists a convex set $\mathcal{P} \subset \Delta(\Omega)$, compact in the weak-convergence topology, such that

$$U(f) = \inf_{\mathcal{P}} Pf = \inf_{\mathcal{P}} \int f dP, \quad f \in \mathcal{F}. \tag{2.1}$$

In Section 7, we relate this specification to the Gilboa and Schmeidler [19] formulation - ours is a specialization - and we provide behavioral foundations for (2.1). Since we suspect that some readers will consider this material to be largely "technical", we defer it to the end.

One difference that may seem important, but that is in fact of minor significance, can be dealt with here. In this paper, acts are taken to be real-valued and they enter linearly into the utility calculation in (2.1). In contrast, Gilboa and Schmeidler and much

[2] The tail σ-algebra is defined by $\Sigma^{tail} = \cap_{t=1}^{\infty} \sigma\left(\vee_{j=t}^{\infty} \mathcal{S}_j\right)$.

[3] In fact, we have overstated the problem somewhat in as much as our central axioms concern only the ranking of acts that depend on finitely many experiments.

of the related literature consider Anscombe-Aumann acts f that have lotteries in $\Delta(Z)$ over a primitive set Z as outcomes. However, this specification of objects of choice can be reduced to ours as follows: Suppose also that there exist best and worst outcomes $\overline{z}$ and $\underline{z}$. Then, under weak conditions, for each state ω and act f, there exists a unique probability p, so that the constant act $f(\omega)$ is indifferent to the lottery $(\overline{z}, p; \underline{z}, 1-p)$; refer to such a lottery as (a bet on) the toss of an (objective) *p-coin*.[4] One can define $u(f(\omega))$ to be this unique probability, so that

$$f(\omega) \sim (\overline{z}, u(f(\omega)); \underline{z}, 1 - u(f(\omega))). \tag{2.2}$$

Such calibration renders the util-outcomes of any act observable, and these are the $[0,1]$-valued outcomes we assume herein and that justify writing utility as in (2.1). A further consequence given (2.1) is that the utility $U(f)$ is also scaled in probability units - it satisfies

$$f \sim (\overline{z}, U(f); \underline{z}, 1 - U(f)). \tag{2.3}$$

Thus f is indifferent to betting on the toss of a $U(f)$-coin.

The fact that *outcomes are "equivalent" probabilities* will be important below - multiplying outcomes, which may seem unnatural, will amount to the very natural operation of multiplying probabilities.

We conclude this section with an elementary lemma that we use repeatedly. Say that $P \in \mathcal{P}$ is a *minimizing,* or *supporting, measure for* f if the infimum in (2.1) is achieved at P. If f is (lower semi-) continuous, such as if f is finitely-based, then there is a minimizer in $\mathcal{P}$, but not so in general.

Lemma 2.2. *Let $f_i \in \mathcal{F}_{fin}$ and $\alpha_i > 0$, $i = 1, ..., n$, with $\Sigma_{i=1}^n \alpha_i = 1$. Then*

$$U(\Sigma_{i=1}^n \alpha_i f_i) = \Sigma_{i=1}^n \alpha_i U(f_i) \tag{2.4}$$

if and only if every measure supporting $\Sigma_{i=1}^n \alpha_i f_i$ also supports every f_i. In particular, (2.4) implies that, for any $m \leq n$, $\beta_i > 0$, and $\Sigma_{i=1}^m \beta_i = 1$,

$$U(\Sigma_{i=1}^m \beta_i f_i) = \Sigma_{i=1}^m \beta_i U(f_i).$$

Proof. Let P^* support the mixed act. Then

$$U(\Sigma_i \alpha_i f_i) = \Sigma_i \alpha_i P^* f_i > \Sigma_i \alpha_i U(f_i),$$

if P^* is not minimizing for some f_i. The rest of the proof is obvious. ■

3. STRONG EXCHANGEABILITY

Turn finally to the core question - how to model the distinction described in the title.

The first part is obvious - if evidence is symmetric, then it is intuitive that an individual would satisfy:

[4] We will not always repeat "objective" below, but there should be no confusion between the motivating coin-tossing experiment described in the introduction, where uncertainty is subjective, and these tosses of an objective coin that define lotteries used to calibrate utility outcomes.

Axiom 1 (SYMMETRY). *For all finitely-based acts f and permutations π, $f \sim \pi f$.*

Assuming subjective expected utility, Symmetry is equivalent to exchangeability of the prior, as noted above. But Symmetry in itself is a relatively weak assumption following, for example, from symmetry of information about all the experiments. The force of the assumption of Symmetry, as reflected in de Finetti's theorem, stems largely from the added assumption of expected utility theory, or a single prior, as will be evident shortly. In a multiple-priors framework, relatively little structure is implied for the set of priors. (See Section 7.1 for a proof.)

Proposition 3.1. *Let $\succeq$ be represented by multiple-priors utility as in (2.1), with set of priors $\mathcal{P}$. Then $\succeq$ satisfies Symmetry iff for every finite permutation π,*

$$P \in \mathcal{P} \implies \pi P \in \mathcal{P}. \tag{3.1}$$

Say that $\mathcal{P}$ is *symmetric* if it satisfies (3.1).

The heart of the paper concerns modeling the perception of "limited evidence of symmetry." Before arguing that the Independence Axiom excludes it, we state the axiom:

INDEPENDENCE: For all α in $(0,1)$,

$$f \succeq g \iff \alpha f + (1-\alpha)\, h \succeq \alpha g + (1-\alpha)\, h.$$

Consider bets in the coin-tossing example. Symmetry implies the indifference

$$H_1T_2 \sim T_1H_2.$$

Here H_1T_2 is the bet that pays 1 util if the first toss yields Heads and the second Tails; the bet T_1H_2 is interpreted similarly. Consider now the choice between either of the above bets and the mixture $\frac{1}{2}H_1T_2 + \frac{1}{2}T_1H_2$, the bet paying $\frac{1}{2}$ if $\{H_1T_2, T_1H_2\}$ and 0 otherwise. The Independence Axiom would imply that

$$\tfrac{1}{2}H_1T_2 + \tfrac{1}{2}T_1H_2 \sim H_1T_2 \sim T_1H_2.$$

This is intuitive given certainty that tossing technique does not vary, since then there is nothing to be gained by mixing; neither is there a cost because outcomes are denominated in utils. On the other hand, if the individual admits the *possibility* that technique varies, and hence that experiments are not identical, then she may strictly prefer the mixture because the bets H_1T_2 and T_1H_2 hedge one another: the former pays well if the first toss is biased towards Heads and the second towards Tails, pays poorly if the opposite bias pattern is valid, and these "good" and "bad" scenarios are reversed for act T_1H_2. Thus $\frac{1}{2}H_1T_2 + \frac{1}{2}T_1H_2$ hedges uncertainty about the bias pattern, and as such, suggests the ranking

$$\tfrac{1}{2}H_1T_2 + \tfrac{1}{2}T_1H_2 \succ H_1T_2 \sim T_1H_2, \tag{3.2}$$

contrary to the Independence Axiom.[5]

[5] Gilboa and Schmeidler [19] suggest that since randomization smooths out payoffs across ambiguous states, a strict preference for randomization reveals an aversion to ambiguity. We rely heavily on similar intuition.

It merits emphasis that concern with the coins not being identical is *not a (probabilistic) risk* - if it were, then, because payoffs are in utils, there would be no value to hedging the risk and hence to randomization. Put another way, it is not possible to model the noted concern by using a single probability measure, since symmetry of information suggests immediately that the associated probability measure is exchangeable, leaving no room for possible differences between coins. This is the heart of Walley's criticism of the exchangeable Bayesian model.

There is another motivation for randomizing which is not derived from the concern that experiments may not be identical. Thus, for example, consider the rankings

$$\tfrac{1}{2}H_1 + \tfrac{1}{2}T_2 \succ H_1 \sim T_2. \tag{3.3}$$

Here we assume for simplicity that Heads and Tails are thought to be equally likely. Suppose further that tossing technique is thought to be irrelevant. Nevertheless, the mixed bet $\frac{1}{2}H_1 + \frac{1}{2}T_2$ may be strictly preferable if there is ambiguity about the physical bias of the coin - this is the key intuition in Gilboa and Schmeidler [19].

Both reasons for randomizing, and hence both forms of violations of Independence, seem important. We do not have a single model that accommodates both (see, however, the remark at the end of Section 5.2). In this paper, we describe two models, each of which accommodates one of (3.2) and (3.3) but not the other.

The next axiom permits only the second rationale for randomizing. Note that it is redundant in the Bayesian case because it is implied by Symmetry and Independence.

STRONG EXCHANGEABILITY: For all finitely-based acts f and all α in $[0,1]$,

$$\alpha f + (1-\alpha)\,\pi f \sim f.$$

Theorem 3.2. *Let $\succeq$ be represented by a multiple-priors utility function as in (2.1), with set of priors $\mathcal{P}$. Then the following conditions are equivalent:*
(i) $\succeq$ satisfies Strong Exchangeability.
(ii) Every prior P in $\mathcal{P}$ is exchangeable.
(iii) There exists $\mathcal{M} \subset \Delta(\Delta(S))$ such that

$$\mathcal{P} = \left\{ \int \ell^{\infty}(\cdot)\, d\mu(\ell) : \mu \in \mathcal{M} \right\}. \tag{3.4}$$

Proof. The equivalence of (ii) and (iii) follows from de Finetti's Theorem.
(ii) $\Longrightarrow$(i): By assumption, for every P in $\mathcal{P}$, act f and permutation π,

$$P(\pi f) = (\pi P)\, f = Pf.$$

For any finitely-based act f, $U(f) = \inf_{P\in\mathcal{P}} Pf = P^{*}f$ for some P^{*} in $\mathcal{P}$. Then

$$U(\pi f) = \inf_{P\in\mathcal{P}} P(\pi f) = \inf_{P\in\mathcal{P}} (\pi P) f = \inf_{P\in\mathcal{P}} Pf = P^{*} f,$$

that is, P^{*} is also minimizing for πf. Therefore, Lemma 2.2 gives the result.

(i)$\Longrightarrow$(ii): Assume Strong Exchangeability. The indifference asserted in the axiom extends to all (not necessarily finitely-based) acts (see Section 7.1). Refer to P^* in $\mathcal{P}$ as an *exposed point* if there exists a continuous act f such that

$$\{P^*\} = \arg\min_{P \in \mathcal{P}} Pf.$$

Then, $\alpha f + (1-\alpha)\,\pi f \sim f \Longrightarrow$ there is a common minimizing measure for f and $\pi f \Longrightarrow P^* = \pi P^*$, and this is true for every π. That is, P^* is exchangeable.

Argue next that $\mathcal{P}$ equals the closed convex hull of its exposed points: Let $c\,(\Omega)$ be the linear space generated by $\Delta\,(\Omega)$; it is separable when endowed with the weak-convergence topology. Therefore, $C\,(\Omega)$, the Banach space of continuous real-valued functions with the sup norm, is an Asplund space [30, Theorem 2.12]. The assertion now follows from Phelps [30, Theorem 5.12].

Finally, (ii) is implied by the fact that the set of all exchangeable measures in $\mathcal{P}$ is closed and convex. (Convexity is obvious. P is exchangeable if *and only if*, for every π and for every $f \in \mathcal{F}_{fin}$,

$$Pf = P\,(\pi f)\,.$$

Since $f \in \mathcal{F}_{fin}$ is continuous, this equality is preserved in the weak-convergence limit.) ■

Part (iii) clarifies how a model with Strong Exchangeability differs from the de Finetti model. Confirming the intuition described preceding the axiom, the representation (3.4) suggests the interpretation whereby the individual is uncertain ex ante which likelihood function applies, but she is certain that the same likelihood function applies to all experiments. This is just as for the Bayesian case - experiments are perceived as identical. The difference here is that the ex ante uncertainty is in general not representable by a single probability measure - there is ambiguity rather than risk regarding the true likelihood function.

Remark 1. *Walley [37, Ch. 9] defines and discusses exchangeability for "previsions" ν, where $\nu\,(f)$ is interpreted as the maximum price (in utils) the individual would be willing to pay for the act f, that is, so that the act $f - \nu\,(f)$ is just desirable. Symmetry of evidence is expressed through indifference between an act f and any permutation πf, in the sense that $\nu\,(f) = \nu\,(\pi f)$. Walley suggests an additional axiom, which he calls exchangeability, which states that*

$$\nu\,(\pi f - f) = 0, \text{ for all } f \text{ and } \pi.$$

The axiom, and his representation result, bear some similarity to Strong Exchangeability and Theorem 3.2. His formulation leads to results that follow almost by definition - for example, the heavy machinery invoked in the proof of our theorem is not needed. Further results for lower previsions appear in de Cooman and Miranda [9] and de Cooman et al [10].

4. NONIDENTICAL EXPERIMENTS

In this section, we describe a model that accommodates the strict preference for randomization in (3.2) and that accordingly, we interpret as capturing a concern that experiments may differ. In terms of the implied representation to be described below (Theorem 5.2), it has in common with de Finetti's (1.1) a single prior, but it differs from his in featuring (in a suitable sense) multiple likelihoods. The model is based on two new axioms, alternatives to Strong Exchangeability.

4.1. Orthogonal Independence

The first axiom is called Orthogonal Independence, and it expresses primarily that poorly understood factors affecting different experiments are unrelated.[6] The point is that randomization *is* a matter of indifference for *some* bets, and precisely when such indifference prevails can be interpreted in terms of the individual's perception of how experiments are related to one another.

Say that the acts f and g *do not hedge one another* if, for every $0 < \alpha < 1$ and p in $[0, 1]$,

$$f \succeq p \Longleftrightarrow [\alpha f + (1-\alpha)\, g \succeq \alpha p + (1-\alpha)\, g]. \tag{4.1}$$

It is easy to see that f and g do not hedge one another if and only if, for all α,

$$U\left(\alpha f + (1-\alpha) g\right) = \alpha U\left(f\right) + (1-\alpha)\, U\left(g\right). \tag{4.2}$$

We use this characterization repeatedly below (without reference).

Think of coin-tossing for concreteness. If tossing techniques are thought to be unrelated across experiments, then presumably the bets H_1 and H_2 do not hedge one another. As pointed out in the discussion of (3.3), bets on different experiments can hedge one another if there is ambiguity about the coin's bias. Here we exclude such ambiguity. Then the unrelatedness of experiments suggests also that H_1T_3 and H_2T_3, for example, do not hedge one another. To illustrate the role of "unrelatedness," suppose that there is concern that outcomes on consecutive tosses could be either perfectly negatively correlated (for example, Heads implies Tails on the next toss) or perfectly positively correlated (Heads implies Heads on the next toss). Then, one would expect the strict preference

$$U\left(\tfrac{1}{2} H_1T_3 + \tfrac{1}{2} H_2T_3\right) > \tfrac{1}{2} U\left(H_1T_3\right) + \tfrac{1}{2} U\left(H_2T_3\right),$$

and hence that H_1T_3 and H_2T_3 hedge one another.

Our axiom builds on this intuition. To express it we generalize "product bets" such as H_1T_3, and consider "product acts." Given any two acts f^* and f, then $f^* \cdot f$ denotes the pointwise product, that is, the act given by

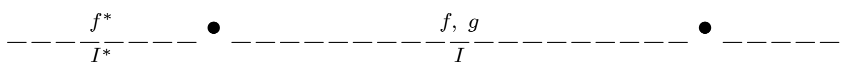
$$\left(f^* \cdot f\right)(\omega) = f^*\left(\omega\right) f\left(\omega\right) \text{ for all } \omega \in \Omega.$$

[6] Since "independence" has a different meaning in an axiomatic context, we often refer to the "unrelatedness" or "stochastic independence" of experiments, though the latter should not be understood in the usual sense of probability theory.

Recall from Section 2.2 that the outcome produced by f in state ω can be viewed as a coin toss which gives the best 'true' underlying outcome $\overline{z}$, or utility 1, with objective probability $f(\omega)$, and the worst outcome $\underline{z}$, or utility 0, with the complementary probability. Similarly, in state ω the product act $f^* \cdot f$ gives a lottery where 1 util is received with objective probability $f^*(\omega) f(\omega)$, corresponding to the independent tosses of the two coins associated with f^* and f.

ORTHOGONAL INDEPENDENCE (OI): If $f, g \in \mathcal{F}_I$ do not hedge one another, then neither do $f^* \cdot f$ and $f^* \cdot g$, for all $f^* \in \mathcal{F}_{I^*}$, with I and I^* finite and disjoint.

The axiom weakens Independence since, by (4.2), nonhedging pairs are precisely those for which utility exhibits the linearity implied by Independence. The reason for the qualifier "Orthogonal" is that one might refer to acts f^* and f as in the statement as being *orthogonal* because they depend on different experiments. Formally, say that f^* and f are (mutually) *orthogonal*, written $f^* \perp f$, if $f^* \in \mathcal{F}_{I^*}$ and $f \in \mathcal{F}_I$ for some disjoint I^* and I. The diagram below illustrates the orthogonality assumed in the axiom. The positioning of acts above the line indicates that f and g depend only on experiments in I, and f^* depends only on those in I^*.



Note that all the acts in the axiom statement are finitely-based.

We will use the following lemma repeatedly, when invoking OI. It illustrates further how Orthogonal Independence, given also multiple-priors utility, expresses the unrelatedness of experiments.

Lemma 4.1. *Let $\succeq$ be represented by a multiple-priors utility function U and satisfy Orthogonal Independence. Then, for all finitely-based acts $f^* \perp f$ and $g^* \perp g$:*

i) f^* and f are nonhedging.

ii) $f^* \cdot f$ and f^* are nonhedging.

iii) If f and g are nonhedging, and if f^* and g^* are either *nonhedging,* or *orthogonal, then $f^* \cdot f$ and $g^* \cdot g$, are nonhedging.*

By (i), acts that depend on different experiments are nonhedging. The remaining parts specify conditions under which nonhedging prevails even where acts depend on overlapping sets of experiments. Some illustrations of nonhedging pairs in the coin-tossing context were provided above. Other examples of such pairs include $\{H_1T_2, T_2\}$, $\{H_1T_3, H_2H_4\}$, and $\{H_1, T_2\}$. The latter case implies, contrary to (3.3), (and assuming again for simplicity that $H_1 \sim T_2$), that

$$\tfrac{1}{2}H_1 + \tfrac{1}{2}T_2 \sim H_1 \sim T_2.$$

Thus, in light of the discussion surrounding (3.3), Orthogonal Independence excludes ambiguity about the coin's bias. However, it does permit (3.2) and thus the concern that experiments may not be identical.

Proof. ii): Since $\succeq$ is a multiple-priors preference, f and the constant act 1 are nonhedging. Thus, ii) follows by OI.

i): Use ii) to derive

$$\begin{aligned} U\left(\tfrac{1}{4}f^* \cdot f + \tfrac{1}{4}f^* + \tfrac{1}{4}f + \tfrac{1}{4}\right) &= U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}\right)\left(\tfrac{1}{2}f + \tfrac{1}{2}\right)\right) \\ &= \tfrac{1}{2}U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}\right) \cdot f\right) + \tfrac{1}{2}U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}\right)\right) \\ &= \tfrac{1}{4}U\left(f^* \cdot f\right) + \tfrac{1}{4}U\left(f^*\right) + \tfrac{1}{4}U\left(f\right) + \tfrac{1}{4}. \end{aligned}$$

By Lemma 2.2 (existence of a common minimizer), i) follows.

iii): Suppose that f^* and g^* are nonhedging. Then, by OI,

$$\begin{aligned} & U\left(\tfrac{1}{4}f^* \cdot f + \tfrac{1}{4}f^* \cdot g + \tfrac{1}{4}g^* \cdot f + \tfrac{1}{4}g^* \cdot g\right) \\ = \; & U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}g^*\right)\left(\tfrac{1}{2}f + \tfrac{1}{2}g\right)\right) = \tfrac{1}{2}U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}g^*\right) \cdot f\right) + \tfrac{1}{2}U\left(\left(\tfrac{1}{2}f^* + \tfrac{1}{2}g^*\right) \cdot g\right) \\ = \; & \tfrac{1}{4}U\left(f^* \cdot f\right) + \tfrac{1}{4}U\left(f^* \cdot g\right) + \tfrac{1}{4}U\left(g^* \cdot f\right) + \tfrac{1}{4}U\left(g^* \cdot g\right) \end{aligned}$$

Apply Lemma 2.2 to conclude that $f^* \cdot f$ and $g^* \cdot g$ are nonhedging.

The case where f^* and g^* are orthogonal is straightforward by the preceding and i). ■

We provide two examples to illustrate what is excluded by Orthogonal Independence.

Example 4.2. *Let P^0 be any countably additive (not necessarily exchangeable) measure and define $\mathcal{P}$ to be the closed convex hull of $\{\pi P^0 : \pi \in \Pi\}$. By construction, $\mathcal{P}$ is symmetric. However, it violates OI.*

There is a simple interpretation: P^0 reflects some asymmetries across experiments, for example, it might be believed that toss 1 is biased towards Heads and that the others are unbiased. If beliefs are instead that there exists exactly one biased toss, though its identity is completely unknown, one is led to $\{\pi P^0 : \pi \in \Pi\}$.[7] Then the agent would be indifferent between betting on Tails for any two coins, but, contrary to OI, she would strictly prefer to randomize, that is,

$$\tfrac{1}{2}T_1 + \tfrac{1}{2}T_2 \succ T_1 \sim T_2.$$

Since the worst case scenario for T_1 (T_2) is that the first (second) coin is the biased one, the mixture smooths out these uncertainties and guarantees at least one coin that is not biased against Tails. Hence it is strictly preferable. OI is violated because the poorly understood factor - which toss is the biased one - relates *the outcomes of the different experiments since there is certainty that only one is biased.*

[7] Taking the closed convex hull has no consequence for decisions.

Example 4.3. *Fix a probability measure ℓ^* in $\Delta(S)$ and let*

$$\mathcal{P} = \{P \in \Delta(\Omega) : mrg_{S_i} P = \ell^* \text{ for all } i\ \}.$$

Thus $\mathcal{P}$ consists of all measures that agree with ℓ^* on each S_i, with joint distributions across different experiments being unrestricted. The interpretation is that there is no ambiguity about the nature of any single experiment, but there is complete ignorance about how experiments are correlated. This perception of the experiments is not covered by our model. The individual in our model is uncertain that experiments are identical because she views each experiment as being affected also by poorly understood factors that vary across experiments, but she is certain that these are unrelated across experiments. Here, in contrast, she is concerned with the possible correlation of these factors across experiments.[8]

Though $\mathcal{P}$ is obviously symmetric, compact and convex, it lies outside the scope of our model because it violates OI as we now show.

For concreteness, let $S = \{H, T\}$ and let ℓ^* describe an unbiased coin toss. OI would imply that

$$\begin{aligned} & U\left(\left(\tfrac{1}{2}H_1 + \tfrac{1}{2}\right)\left(\tfrac{1}{2}T_2 + \tfrac{1}{2}\right)\left(\tfrac{1}{2}H_3 + \tfrac{1}{2}\right)\right) \\ = \ & \tfrac{1}{2}U\left(\left(\tfrac{1}{2}H_1 + \tfrac{1}{2}\right)\left(\tfrac{1}{2}T_2 + \tfrac{1}{2}\right)H_3\right) + \tfrac{1}{2}U\left(\left(\tfrac{1}{2}H_1 + \tfrac{1}{2}\right)\left(\tfrac{1}{2}T_2 + \tfrac{1}{2}\right)\right) \\ & \ldots \\ = \ & \tfrac{1}{8}\left[\begin{array}{c} U(1) + U(H_1) + U(T_2) + U(H_3) \\ +U(H_1 \cdot T_2) + U(T_2 \cdot H_3) + U(H_1 \cdot H_3) + U(H_1 \cdot T_2 \cdot H_3). \end{array}\right]. \end{aligned}$$

Thus, there is a common minimizing measure, say P, for the acts $H_1, T_2, H_3, H_1 \cdot T_2, T_2 \cdot H_3, H_1 \cdot H_3$ and $H_1 \cdot T_2 \cdot H_3$. Compute that

$$\begin{aligned} U(H_1) &= U(T_2) = U(H_3) = \tfrac{1}{2}, \text{ and} \\ U(H_1 \cdot T_2) &= U(T_2 \cdot H_3) = U(H_1 \cdot H_3) = U(H_1 \cdot T_2 \cdot H_3) = 0, \end{aligned}$$

where, for example, $U(H_1 \cdot T_2) = 0$ because the worst-case scenario for this act is that tosses 1 and 2 are perfectly positively correlated. Since P is a common minimizer, deduce that

$$\begin{aligned} P(H_1) &= P(T_2) = P(H_3) = \tfrac{1}{2}, \text{ and} \\ P(H_1T_2) &= P(T_2H_3) = P(H_1H_3) = P(H_1T_2H_3) = 0. \end{aligned}$$

But there does not exist a probability measure satisfying these conditions. (Since $P(H_1T_2) = 0$, $P(H_1H_3) = 0$ and $P(H_1) = \frac{1}{2}$, it follows that

$$\begin{aligned} P(H_1H_2H_3) &= 0, \ P(H_1H_2T_3) = \tfrac{1}{2}, \\ P(H_1T_2H_3) &= 0 \text{ and } P(H_1T_2T_3) = 0. \end{aligned}$$

Combine these with $P(T_2) = \frac{1}{2}$ and $P(T_2H_3) = 0$ to deduce that

$$P(T_1T_2H_3) = 0 \text{ and } P(T_1T_2T_3) = \tfrac{1}{2}.$$

Finally, use $P(H_3) = \frac{1}{2}$ to conclude that $P(T_1H_2H_3) = \frac{1}{2}$. But then $P(H_1H_2T_3) + P(T_1T_2T_3) + P(T_1H_2H_3) > 1$.)

[8] In fact, the difference is more subtle, since, as shown in the sequel, OI does permit the perception of some degree of dependence between experiments.

4.2. A Final Axiom: Super-Convexity

Denote by θ the *shift* operator, so that, for any act,

$$(\theta f)(s_1, s_2, s_3, ...) = f(s_2, s_3, ..);$$

θ^n denotes the n-fold replication of θ. It is straightforward to show that Symmetry implies also indifference to shifts,[9]

$$\theta f \sim f \quad \text{for all } f \in \mathcal{F}.$$

For any act $g^* \in \mathcal{F}_{\{1,...,n\}}$, the acts g^* and $\theta^n f$ are orthogonal, and their product is given by

$$(g^* \cdot \theta^n f)(\omega) = g^*(s_1, ..., s_n) f(s_{n+1}, s_{n+2}, ...).$$

The final axiom strengthens the assumption of convexity of preference, one of the central axioms in Gilboa and Schmeidler's [19] characterization of multiple-priors (following Schmeidler [35], they refer to it as uncertainty, or ambiguity, aversion). Convexity has the standard meaning that sets of the form $\{f \in \mathcal{F} : f \succeq g\}$ are convex. Given also the other axioms, (notably Certainty Independence), used to characterize the multiple-priors model, the preceding convexity is equivalent to concavity of the utility function U that represents preference via probability equivalents as in (2.3). For this reason we call our stronger assumption Super-Convexity.

SUPER-CONVEXITY: Let U be the probability-equivalent utility function (as in (2.3)) representing the preference $\succeq$. Then, for all $g^*, h^* \in \mathcal{F}_{\{1,...,n\}}$, with g^* and h^* nonhedging, and $g^* \geq h^*$, the function $W : \mathcal{F}_{fin} \to \mathbb{R}$ defined by

$$W(f) = U(g^* \cdot \theta^n f) - U(h^* \cdot \theta^n f),$$

is concave.

In the special case $g^* = 1$ and $h^* = 0$, the axiom imposes concavity of $U(\cdot)$ on $\mathcal{F}_{fin}$, as in the Gilboa-Schmeidler model. We emphasize that Super-Convexity is an assumption about preference: since the utility function U gives the probability equivalents of acts, the axiom can be expressed explicitly and exclusively in terms of preference.

Finally, it can be understood as follows. For any F' and F, acts over experiments beyond the n^{th}, because of hedging gains, the individual prefers the mixed act $\alpha F' + (1-\alpha)F$ as expressed by the concavity of $U(\cdot)$. The same is true if the acts and the mixed act are premultiplied by $g^* \in \mathcal{F}_{\{1,...,n\}}$, or by $h^* \in \mathcal{F}_{\{1,...,n\}}$. However, the value of mixing is small if h^* is "small" at every state, since then premultiplication by h^* shrinks differences between F', F and $\alpha F' + (1-\alpha)F$. (In the extreme case where $h^* = 0$, compounding by h^* wipes out all differences between acts.) For this reason mixing has greater value when premultiplication is by g^*, $g^* \geq h^*$. The restriction that g^* and h^* be nonhedging weakens the axiom; in fact, the stronger axiom without that restriction is implied given the other axioms, as can be seen from the representation derived below.

[9] By Symmetry, $U(\theta f) = U(f)$ on $\mathcal{F}_{fin}$. By Epstein and Wang [17, Theorem D.2] and Lemma B.8, the two functions coincide everywhere.

Though we believed, at an earlier stage in this research, that Super-Convexity was implied by the other axioms, that remains an open question. On the other hand, Super-Convexity does not imply Orthogonal Independence, even given the other axioms, as illustrated by the utility function (5.7) described below.

5. REPRESENTATIONS: "CONDITIONALLY IID"

5.1. A Definition

Our next objective is to describe the representation implied by Symmetry, Orthogonal Independence and Super-Convexity. It is our counterpart, or generalization, of the "conditionally i.i.d." representation in de Finetti's theorem. Thus we begin with a definition of "stochastic independence" of experiments for our framework. (Here we mean that experiments are unrelated, not even by a common bias in the case of coin tossing - think of the case where the bias is known with certainty.) In the Bayesian setting, it amounts to beliefs being represented by a product measure. However, the situation is more complicated in a multiple-priors framework - there are different ways of defining a "product" set of priors consistent with given sets of marginals (for example, see Hendon et al [22] and Ghirardato [18]).

We define "product" in terms of utility functions rather than directly in terms of sets of priors. Say that the multiple-priors utility function U, as in (2.1), is a *product utility function* if

$$U(f \cdot g) = U(f)\, U(g) \text{ for all orthogonal } f, g \in \mathcal{F}_{fin}. \tag{5.1}$$

If also preference represented by U satisfies Symmetry, then refer to an *IID utility function*, and to the corresponding set of priors $\mathcal{P}$ as an *IID set of priors.* Following [15], the acronym IID stands for "independently and *indistinguishably* (as opposed to identically) distributed."[10]

The rationale for (5.1) may seem obvious, but its behavioral meaning should be made clear. Recall the probability-equivalence nature of outcomes and utility (see (2.2) and (2.3)). The acts f and g are assumed to depend on different experiments - for concreteness, let $f \in \mathcal{F}_1$ and $g \in \mathcal{F}_2$. Then, in state (s_1, s_2), $f \cdot g$ yields (the equivalent of) successive and independent tosses of an objective $f(s_1)$-coin and an objective $g(s_2)$-coin.[11] If experiments 1 and 2 are "stochastically independent," it is intuitive to perceive this prospect as though the "order" of coin tossing were: toss all $f(s_1)$-coins as s_1 varies over S_1, and separately and *independently* toss all $g(s_2)$-coins as s_2 varies over S_2. But the prospect consisting of the first set of coin tosses is equivalent to f, and the second to g. Further, f is indifferent to a $U(f)$-coin and g is indifferent to a $U(g)$-coin. Conclude that $f \cdot g$ is indifferent to winning 1 util if both the $U(f)$- and the $U(g)$-coins, tossed independently, produce favorable outcomes, which is equivalent to a $U(f)\, U(g)$-coin. This "proves" that (5.1) is implied if experiments are seen to be independent. The converse is similarly intuitive.

[10] We continue to use the lower case acronym iid when referring to single measures, with the usual meaning of "independently and identically distributed."

[11] A p-coin is one that yields 1 util (or the best outcome $\overline{z}$) with objective probability p and 0 utils (or the worst outcome $\underline{z}$) with probability $1 - p$.

Lemma 5.1. *If U is an IID utility function, then U satisfies both Orthogonal Independence and Super-Convexity.*

Proof. Take f, g and f^* as in the statement of Orthogonal Independence. Then,

$$\begin{aligned} & U\left(\tfrac{1}{2}f^* \cdot f + \tfrac{1}{2}f^* \cdot g\right) = U\left(f^* \cdot \left(\tfrac{1}{2}f + \tfrac{1}{2}g\right)\right) \\ = \; & U\left(f^*\right)\left(\tfrac{1}{2}U\left(f\right) + \tfrac{1}{2}U\left(g\right)\right) = \tfrac{1}{2}U\left(f^* \cdot f\right) + \tfrac{1}{2}U\left(f^* \cdot g\right). \end{aligned}$$

Thus $f^* \cdot f$ and $f^* \cdot g$ are nonhedging by (4.2).

Super-Convexity follows from the fact that

$$\begin{aligned} U\left(g^* \cdot \theta^n f\right) - U\left(h^* \cdot \theta^n f\right) &= \left[U\left(g^*\right) - U\left(h^*\right)\right] U\left(\theta^n f\right) \\ &= \left[U\left(g^*\right) - U\left(h^*\right)\right] U\left(f\right), \end{aligned}$$

and $U\left(g^*\right) \geq U\left(h^*\right)$ if $g^* \geq h^*$. ■

To help fix ideas, we describe one example of an IID utility. Fix a (closed) set $\mathcal{L}$ of probability measures on S, thought of as the set of priors applying to any single experiment. Let[12]

$$\mathcal{P}_{WF} = clh\left(\mathcal{L}^\infty\right), \text{ where } \mathcal{L}^\infty \equiv \left\{\otimes_{i\in\mathbb{N}}\ell_i : \ell_i \in \mathcal{L} \text{ for every } i\right\}. \tag{5.2}$$

Since the utility of any finitely-based act is a minimum over $\mathcal{L}^\infty$, which consists exclusively of product measures, (5.1) is obvious; so is Symmetry. Therefore, U_{WF} defined by

$$U_{WF}\left(f\right) = \inf_{P\in\mathcal{L}^\infty} Pf, \;\; f \in \mathcal{F}, \tag{5.3}$$

is an IID utility function. This product is adapted from Walley and Fine [38], and has been studied also by Gilboa and Schmeidler [19].

We emphasize that U_{WF} is just one example of an IID utility function. It is well-known in the decision theory literature (see Hendon *et al* [22] and Ghirardato [18]) that stochastic independence is multi-faceted in the multiple-priors (or nonadditive probability) framework, and hence that there is more than one way to form an independent product from a given set $\mathcal{L}$ of priors over S. In other words, in general, and in contrast to the Bayesian setting, there are many utility functions satisfying (5.1), and hence the "stochastic independence" embodied in it, that also agree on the ranking of acts over any single experiment.

5.2. A Representation Result

Some preliminaries are needed in order to state the representation. Any set of priors $\mathcal{P}$ lies in $\mathcal{K}^c\left(\Delta\left(\Omega\right)\right)$, the space of compact and convex subsets of $\Delta\left(\Omega\right)$; the Hausdorff metric topology renders it compact metric.

[12] $\otimes_{t\in I}\ell_t$ denotes the unique countably additive product measure with marginals ℓ_t. Since $\mathcal{L}^\infty$ is not convex, we take its closed convex hull, denoted by $clh\left(\mathcal{L}^\infty\right)$, in order to conform to the normalization that sets of priors be closed and convex. The sets $clh\left(\mathcal{L}^\infty\right)$ and $\mathcal{L}^\infty$ generate the identical preference.

Each $\mathcal{P} \in \mathcal{K}^c(\Delta(\Omega))$ corresponds to a unique multiple-priors preference, or equivalently, to a unique multiple-priors utility function $U_{\mathcal{P}} : \mathcal{F} \to \mathbb{R}$, given by

$$U_{\mathcal{P}}(f) = \inf_{P \in \mathcal{P}} Pf.$$

This correspondence induces a compact metric topology on

$$\mathcal{U} = \{U_{\mathcal{P}} : \mathcal{P} \in \mathcal{K}^c(\Delta(\Omega))\}.$$

The subset of IID utility functions,

$$\mathcal{V} = \{U \in \mathcal{U} : U \text{ is IID}\},$$

inherits the induced topology.

Theorem 5.2. *The preference $\succeq$ on $\mathcal{F}$ is a multiple-priors preference and satisfies Symmetry, Orthogonal Independence and Super-Convexity if and only if it admits representation by a utility function U of the form in (2.1) satisfying*

$$U(f) = \int_{\mathcal{V}} V(f)\, d\mu(V), \text{ for all } f \text{ in } \mathcal{F}, \tag{5.4}$$

for some Borel probability measure μ on $\mathcal{V}$. Moreover, μ is unique.

The proof of sufficiency is relegated to Appendix B. Here consider briefly necessity (see Appendix B for further details). The first step is to verify that the integrand on the right is well-defined for every f. This is done by showing that the function $V \longmapsto V(f)$ is universally measurable, and by making use of the fact that any measure μ admits a unique extension, also denoted μ, to the universal completion of Σ. (A similar procedure is used throughout, without explicit mention, to make sense of integrals where measurability issues arise.)

Turn to axioms. Since U is a mixture of symmetric utility functions, it is also symmetric. We showed above (Lemma 5.1) that Orthogonal Independence and Super-Convexity are satisfied by any IID utility function - the argument is readily extended to any mixture of IID utility functions as in the representation.

The theorem generalizes de Finetti's, wherein each IID utility function in the support of μ is an expected utility function with i.i.d. probabilistic beliefs. The more general representation (5.4) suggests an interpretation similar to that familiar for a mixture of i.i.d. beliefs. Any IID utility function reflects the view that experiments are indistinguishable (because of Symmetry) and unrelated or independent. Thus experiments would be IID (indistinguishable and independent) if the individual knew which IID utility function were appropriate or correct. However, she is uncertain of that, as reflected by the measure μ. Overall, therefore, she views experiments as being IID conditionally on the correct V. Because the possible functions V correspond to multiple-priors rather than to expected utility, the individual may value randomization, as illustrated in (3.2), and accordingly not view experiments as being identical.

To illustrate, suppose that each IID function in the support of μ has the form in (5.3). Then, in a slight abuse of notation, where the uncertainty modeled by μ is translated into uncertainty about the true set $\mathcal{L}$,

$$U(f) = \int \left(\inf_{\mathcal{L}^\infty} Pf \right) d\mu(\mathcal{L}). \tag{5.5}$$

Given resolution of that uncertainty and thus a specific $\mathcal{L}$, the same set is assumed to describe each experiment (because the minimum is over $\mathcal{L}^\infty$). This implies that experiments are indistinguishable (or viewed symmetrically). However, experiments are not viewed as identical because $\mathcal{L}^\infty$ admits that different likelihoods from $\mathcal{L}$ apply to different experiments.

In the concrete setting of coin-tossing, any (convex) set $\mathcal{L}$ of likelihoods can be identified with an interval $\mathcal{I} = [\mathcal{I}_m, \mathcal{I}_M] \subset [0,1]$, interpreted as a set of possible probabilities for Heads. There is ex ante uncertainty about which interval is the correct one, but conditional on knowing $\mathcal{I}$, coin tosses are viewed as indistinguishable ambiguous experiments, independent from one another in the specific sense of (5.2). It is easily seen how the model can accommodate the value of randomization in (3.2), and this is so even when there is certainty about $\mathcal{I}$. Suppose $\mathcal{I}_m < \frac{1}{2} < \mathcal{I}_M$. Then

$$\begin{aligned} U\left(\tfrac{1}{2}H_1T_2 + \tfrac{1}{2}T_1H_2\right) &= \min_{P \in \mathcal{L}^2} P\left(\tfrac{1}{2}H_1T_2 + \tfrac{1}{2}T_1H_2\right) \\ &= \min_{\ell_1,\ell_2 \in \mathcal{L}} \tfrac{1}{2}\left(\ell_1(H_1)\ell_2(T_2) + \tfrac{1}{2}\ell_1(T_1)\ell_2(H_2)\right) \\ &= \min_{\ell_1,\ell_2 \in \mathcal{L}} \tfrac{1}{2}\left[\ell_1(H_1)(1-\ell_2(H_2)) + (1-\ell_1(H_1))\ell_2(H_2)\right] \\ &= \min\left\{\mathcal{I}_m(1-\mathcal{I}_m), \mathcal{I}_M(1-\mathcal{I}_M)\right\} \\ &\geq \mathcal{I}_m(1-\mathcal{I}_M) = U(H_1T_2) = U(T_1H_2). \end{aligned}$$

The representation result leads to an interesting implication about the perceived value of repetition, which, at a mathematical level, extends the fact that for any random sequence (X_t) having an exchangeable probability law, X_i and X_j are positively correlated if $i \neq j$.[13]

Theorem 5.3. *If the multiple-priors preference $\succeq$ satisfies Symmetry, Orthogonal Independence and Super-Convexity, then, for any act $f \in \mathcal{F}_{\{1,\dots,n\}}$,*

$$f \sim p \implies f \cdot \theta^n f \succeq p^2. \tag{5.6}$$

Proof. By the representation,

$$\begin{aligned} U(f \cdot \theta^n f) &= \int V(f \cdot \theta^n f)\, d\mu(V) = \int V(f)\, V(\theta^n f)\, d\mu(V) \\ &= \int (V(f))^2\, d\mu(V) \geq \left(\int V(f)\, d\mu(V) \right)^2 = (U(f))^2 = p^2, \end{aligned}$$

[13] In Hewitt and Savage [23], see their Theorem 5.1.

where we use the fact that every V is symmetric (and hence also invariant to shifts) and a product utility function, and also the familiar property that the geometric average is at least as large as the arithmetic average. ■

For simplicity, consider the special case of bets (binary acts). Suppose that a bet on A is indifferent to the bet on a coin with known objective probability p. How would an individual rank two-fold repetitions of each? In the case of the coin, the two tosses would be independent and thus have probability p^2 of success. For the subjective bet, the repetitions are not plausibly viewed as independent in general, as de Finetti pointed out in the Bayesian setting. Where there is a common element connecting experiments - like the uncertain bias of a coin that is tossed repeatedly - experiments are presumably viewed as "positively correlated", which makes bets such as $A \times A$ more attractive than two-fold independent replicas of the bet on A. This intuition relies only on the individual having a "conditionally i.i.d. (or IID)" view of experiments and not on the experiments conforming to a Bayesian (probabilistic) model.[14]

Remark 2. *A more general functional form that is likely to have occurred to many readers is:*

$$U(f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V(f)\, d\mu, \tag{5.7}$$

where $\mathcal{M} \subset \Delta(\mathcal{V})$ is a set of probability measures over the set $\mathcal{V}$ of IID utility functions. It is not difficult to see that this functional form, with each V being a Walley-Fine IID utility function, for example, can accommodate both of the motivating behaviors (3.2) and (3.3). More generally, it includes both of our models as special cases and seems like an obvious candidate as the missing unifying model. The model satisfies Symmetry and Super-Convexity, but not Orthogonal Independence.[15] We do not have an axiomatic characterization of (5.7).

5.3. The Representation of Sets of Priors

The representation given in Theorem 5.2 is for utility functions, while de Finetti's theorem is about beliefs. The former seems more appropriate for a decision-theoretic model, but it is interesting to consider also a formulation that is closer to de Finetti's. In his theorem, every exchangeable measure is represented as a mixture of i.i.d. measures. Here, every set of priors consistent with our axioms is a (suitably defined) mixture of IID sets of priors.

To state this formally, define

$$\Gamma = \{\mathcal{Q} \in \mathcal{K}^c(\Delta(\Omega)) : \mathcal{Q} \text{ is an IID set}\}.$$

Since Γ is homeomorphic to $\mathcal{V}$, each measure on $\mathcal{V}$ corresponds to a unique measure on Γ, and we use the same symbol to denote both. Given $\mu \in \Delta(\Gamma)$, use Aumann's integral for a correspondence to define the set of priors $\int_\Gamma \mathcal{Q} d\mu(\mathcal{Q})$. (Technical details are provided in Appendix B, which also contains, in Section B.1, all the ingredients of a proof of the following Corollary.)

[14] Even in the Bayesian case, we have not found intuition for (5.6) that relies solely on the axioms, without recourse to the representation.

[15] Details are omitted.

Corollary 5.4. *Let $\succeq$ be represented by multiple-priors utility U as in (2.1) with set of priors $\mathcal{P}$. Then each hypothesis in Theorem 5.2 is equivalent to $\mathcal{P}$ being expressible in the form*[16]

$$\mathcal{P} = cl\left(\int_{\Gamma} \mathcal{Q} d\mu\left(\mathcal{Q}\right)\right),$$

for the Borel probability measure μ on Γ corresponding to the measure on $\mathcal{V}$ appearing in (5.4).

When all IID sets have the form in (5.2), one obtains a representation even closer to de Finetti's. Then, with the obvious abuse of notation,

$$\mathcal{P} = cl\left(\int clh\left(\mathcal{L}^{\infty}\right) d\mu\left(\mathcal{L}\right)\right).$$

De Finetti's representation (1.1) is the special case where there is certainty that each set of likelihoods is a singleton, and hence that each experiment is described by the same likelihood. Here, by contrast, *multiple likelihoods* are associated with each experiment.[17]

5.4. Ambiguity and Dissimilarity

The next theorem describes axiomatically the gap between de Finetti's model and ours.

Theorem 5.5. *Let the multiple priors preference $\succeq$ satisfy Symmetry, Orthogonal Independence and Super-Convexity. Then the following statements are equivalent:*

(i) $\succeq$ is an expected utility preference.

(ii) $\succeq$ satisfies Strong Exchangeability on the subdomain of acts over $S_1 \times S_2$, that is,

$$\alpha f + (1-\alpha)\pi f \sim f \quad \text{for all } f \in \mathcal{F}_{\{1,2\}}.$$

(iii) $\succeq$ satisfies the Independence Axiom on the subdomain $\mathcal{F}_1$ of acts over S_1.

(i) is the de Finetti model. The other conditions describe alternative characterizations of how it differs from ours. According to intuition given earlier, (ii) says that the first two experiments, and hence also any other pair, are perceived as identical. Following Gilboa and Schmeidler, we think of violations of Independence as reflecting (aversion to) ambiguity. Therefore, (iii) says that the first experiment is unambiguous. Conclude that our model permits any *two* experiments to be nonidentical by allowing ambiguity about any *single* experiment. This connection seems to us to be intuitive (however, see Example 4.3, for a specification where it is violated).

Proof. (i)$\Longrightarrow$(ii) : clear.

[16] $cl(\cdot)$ denotes closure. Below $clh(\cdot)$ denotes closed convex hull.

[17] Contrast also with the representation (3.4), corresponding to Strong Exchangeability, where every experiment is described by the same likelihood but where there is ambiguity about which likelihood is the correct one.

(ii)$\Longrightarrow$(iii) : Let U represent $\succeq$ and assume (ii). For $f, g \in \mathcal{F}_1$,

$$\begin{aligned}
& U\left(\tfrac{1}{8}\left[f \cdot \theta g + \theta g + f + 1 + g \cdot \theta f + \theta f + g + 1\right]\right) \\
= & U\left(\tfrac{1}{2}\left(\tfrac{1}{2}f + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta g + \tfrac{1}{2}\right) + \tfrac{1}{2}\left(\tfrac{1}{2}g + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta f + \tfrac{1}{2}\right)\right) \\
= & U\left(\left(\tfrac{1}{2}f + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta g + \tfrac{1}{2}\right)\right) \quad \text{(by (ii))} \\
= & \tfrac{1}{2}U\left(\left(\tfrac{1}{2}f + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta g + \tfrac{1}{2}\right)\right) + \tfrac{1}{2}U\left(\left(\tfrac{1}{2}g + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta f + \tfrac{1}{2}\right)\right) \quad \text{(by Symmetry)} \\
= & \tfrac{1}{8}\left[U\left(f \cdot \theta g\right) + U\left(\theta g\right) + U\left(f\right) + U\left(1\right) + U\left(g \cdot \theta f\right) + U\left(\theta f\right) + U\left(g\right) + U\left(1\right)\right]. \\
& \text{(by Orthogonal Independence)}
\end{aligned}$$

Thus, by Lemma 2.2, there is a common minimizing measure for f and g, and (iii) follows.

(iii)$\Longrightarrow$(i) : By Theorem 5.2, there exists $\mu \in \Delta(\mathcal{V})$ such that $U(f) = \int V(f)\, d\mu(V)$, for all f in $\mathcal{F}$. By (iii),

$$\int \left[V\left(\alpha f + (1-\alpha) g\right) - \alpha V(f) - (1-\alpha) V(g)\right] d\mu(V) = 0 \text{ for all } f, g \in \mathcal{F}_1.$$

Since the integrand is nonnegative for all $V \in \mathcal{V}$, conclude that: for all $f, g \in \mathcal{F}_1$, *a.s.*-$\mu[V]$,

$$V\left(\alpha f + (1-\alpha) g\right) = \alpha V(f) + (1-\alpha) V(g). \tag{5.8}$$

Thus it suffices to show that if $V \in \mathcal{V}$ satisfies (5.8), then V is an expected utility function.

Assume V satisfies (5.8) and let $\mathcal{P}$ be the corresponding set of measures for V. Write $V(B)$ instead of $V(1_B)$. By the assumption, there exists $\ell \in \Delta(S)$ such that, for all $A \in \Sigma_1$,

$$V(A) = \ell(A).$$

Claim: If $P \in \mathcal{P}$, then $P(A_1 \times A_2 \times \cdots \times A_n) = V(A_1 \times A_2 \times \cdots \times A_n)$ for all $A_i \in \Sigma_i$, $i \leq n$.

Let $A = A_1 \times A_2 \times \cdots \times A_n$. Since S is finite,

$$\begin{aligned}
P(A) &= 1 - P(\Omega \backslash A) \leq 1 - V(\Omega \backslash A) \\
&\leq 1 - \sum_{(s_1,\ldots s_n) \notin A} V\left(\{(s_1, \ldots, s_n)\}\right) \\
&= 1 - \sum_{(s_1,\ldots s_n) \notin A} \prod_{i=1}^{n} V\left(\{s_i\}\right) = 1 - \sum_{(s_1,\ldots s_n) \notin A} \prod_{i=1}^{n} \ell\left(\{s_i\}\right) \\
&= \sum_{(s_1,\ldots s_n) \in A} \prod_{i=1}^{n} \ell\left(\{s_i\}\right) = \prod_{i=1}^{n} \ell\left(\{A_i\}\right) = \prod_{i=1}^{n} V\left(\{A_i\}\right) \\
&= V(A).
\end{aligned}$$

But, $P(A) \geq \min_{P' \in \mathcal{P}} P'(A) = V(A)$. Thus, $P(A) = V(A)$.

Conclude that all $P \in \mathcal{P}$ agree with V, and therefore, with one another, on finite rectangles. Since finite rectangles generate the Borel σ-algebra Σ, $\mathcal{P}$ is a singleton. ■

6. UPDATING

There is a given ordering of experiments (which need not be temporal); $s_1^n = (s_1, ..., s_n)$ denotes a generic sample or history of length n. Ex ante preference on $\mathcal{F}$ is $\succeq_0$, and $\succeq_{n,s_1^n}$ denotes preference on $\mathcal{F}$ conditional on the sample s_1^n. (When there is no need to emphasize the sample, we suppress it in the notation and write $\succeq_n$; similarly for other random variables.) We seek a model that describes how preferences evolve along a sample.

There is an implicit assumption in this set up which should be made explicit. We have defined outcomes in terms of util/probability equivalents, which obviously depends on how the individual ranks lotteries (constant acts) over the underlying physical outcomes (represented earlier by the set Z). This rescaling of outcomes is straightforward when dealing with a single preference order. However, when there are several preferences, as is the case here, in general they may disagree on how to rank lotteries, and thus any given physical action would translate into a different act depending on which preference order was being considered. Our implicit assumption is that $\succeq_0$ and every conditional preference $\succeq_n$ agree on the ranking of lotteries. That justifies interpreting any given f in $\mathcal{F}$ as representing the same physical action for all the noted preferences.

Our model of updating applies to the second model above, where experiments are not necessarily identical. Thus assume that $\succeq_0$ and every $\succeq_n$ satisfy the axioms of Theorem 5.2, namely Symmetry, Orthogonal Independence and Super-Convexity. Call this composite axiom *BASIC*.

We assume also Consequentialism - the conditional ranking given the sample s_1^n does not take into account what the acts might have delivered had a different sample been realized. Formally, we assume:

CONSEQUENTIALISM: $f' \sim_{n,s_1^n} f$ if $f'(s_1^n, \cdot) = f(s_1^n, \cdot)$.

6.1. Weak Dynamic Consistency

We postulate the following weak form of dynamic consistency. Abbreviate $\mathcal{F}_{\{n+1,n+2,...\}}$ by $\mathcal{F}_{>n}$.

WEAK DYNAMIC CONSISTENCY (WDC): For any $n \geq 1$, sample s_1^{n-1}, and acts $f', f \in \mathcal{F}_{>n}$,

$$f' \succeq_{n,(s_1^{n-1},s_n)} f \text{ for all } s_n \implies f' \succeq_{n-1,s_1^{n-1}} f, \text{ and}$$

$$f' \succ_{n,(s_1^{n-1},s_n)} f \text{ for some } s_n \implies f' \succ_{n-1,s_1^{n-1}} f.$$

If the defining conditions are assumed to hold for all acts f' and f, then one obtains the usual notion of dynamic consistency that we abbreviate DC. In that case, when the acts f' and f can depend on all experiments, each s_i is *both* a signal *and* a payoff-relevant state. In contrast, for each comparison in WDC, states are *either* signals $(s_1, ..., s_n)$, *or* payoff-relevant $(s_{n+1}, ...)$, but *not both*. Thus WDC requires dynamic consistency in the ranking of terminal payoffs as 'pure signals' are received and beliefs and rankings of future prospects are updated.

Note that WDC is weaker than DC even in the Bayesian context. DC implies Bayes' Rule, but, as will become evident below, WDC does not. On the other hand, as argued in the introduction, it is strong enough to accommodate important settings. There are many cases where an individual observes signals and uses them to learn about a payoff relevant "parameter". Here the signals are $(s_1, ..., s_n)$ for some n, and the parameter is $(s_{n+1}, s_{n+2}, ...)$.

Since all utility functions satisfy the axioms in Theorem 5.2, each admits a representation in terms of a unique measure over $\mathcal{V}$, the set of IID utility functions. Their utility functions are U_0 and $U_n(\cdot \mid s_1^n)$, for $\succeq_0$ and $\succeq_{n,s_1^n}$ respectively; frequently, dependence on the sample is suppressed and we write simply $\succeq_n$ and U_n. Then

$$U_0(f) = \int_{\mathcal{V}} V(f)\, d\mu_0(V), \text{ for all } f \in \mathcal{F},$$

and, imposing Consequentialism,

$$U_n(f \mid s_1^n) = \int_{\mathcal{V}} V(f(s_1^n, \cdot))\, d\mu_n(V), \text{ for all } f \in \mathcal{F},$$

for some probability measure μ_n that depends on the realized sample s_1^n. The updating problem thus reduces to describing the evolution of μ_n as a function of μ_0 and the realized sample.

The implications of WDC and the other axioms are described in terms of a *likelihood function* $L : \mathcal{V} \to \Delta(\Omega)$, where $V \longmapsto L(B \mid V)$ is (Borel) measurable for each measurable subset B of Ω. Think of $L(B \mid V)$ as the likelihood of $B \subset \Omega$, a set of infinite samples, conditional on V describing the perception of experiments. These likelihoods are used in describing inferences drawn after observing a sample; they are not to be thought of as describing ex ante beliefs. For each n and likelihood function L, L_n is its one-step-ahead conditional at stage n, $L_n : S^{n-1} \times \mathcal{V} \to \Delta(S)$.[18] Thus for each sample s_1^{n-1}, $L_n(\cdot \mid V) \in \Delta(S)$ gives the probability distribution, or likelihood, for the n^{th} experiment, conditional on s_1^{n-1} and the given V.

The central result in our model of updating follows.

Theorem 6.1. *The axioms Basic, Consequentialism and WDC are satisfied if and only if the representing probability measures $\{\mu_n\}$ are related as follows: there exists a likelihood function L such that, for all $n \geq 1$,*

$$d\mu_n(V) = \frac{L_n(s_n \mid V)}{\overline{L}_n(s_n)} d\mu_{n-1}(V), \tag{6.1}$$

where

$$\overline{L}_n(\cdot) = \int L_n(\cdot \mid V)\, d\mu_{n-1}(V), \tag{6.2}$$

is a probability measure on S having full support.

[18]More precisely, $L_n(\cdot \mid V)$ is a regular conditional probability on S_n given s_1^n (suppressed in the notation), which exists as long as S is Polish.

The proof of necessity is straightforward. We verify only WDC. For any $f \in \mathcal{F}_{>n}$,

$$\begin{aligned}\Sigma_{s_n} \overline{L}_n(s_n) U_n(f \mid s_n) &= \Sigma_{s_n} \left(\int V(f) L_n(s_n \mid V) d\mu_{n-1} \right) \\ &= \int V(f) \left(\Sigma_{s_n} L_n(s_n \mid V) \right) d\mu_{n-1} \\ &= \int V(f) d\mu_{n-1} = U_{n-1}(f),\end{aligned}$$

or

$$U_{n-1}(f) = \Sigma_{s_n} \overline{L}_n(s_n) U_n(f \mid s_n), \; f \in \mathcal{F}_{>n}, \tag{6.3}$$

which implies WDC.

See Appendix D for the proof of sufficiency. The argument amounts to showing that the problem is a special case of affine aggregation (see de Meyer and Mongin [27]); other special cases include Harsanyi's aggregation theorem [21] and probability aggregation [29].

The theorem may be surprising at first glance and some discussion is in order. Two features stand out: (i) likelihood functions are not tied to the ex ante preference $\succeq$; and (ii) the implied process of posteriors $\{\mu_n\}$ is identical to that implied by a suitable Bayesian model. We elaborate on each in turn.

In the absence of ambiguity, when prior beliefs are probabilistic, it is standard practice to use them to define likelihood functions for updating, as in Bayes' Rule. The normative argument for doing so is that Bayesian updating delivers DC. However, if only WDC is sought, then even under subjective expected utility, one can use any likelihood function to define updating. Also more generally, *any* likelihood function L can be used for updating in such a way as to satisfy WDC. In particular, though L is derived from the entire set of (conditional) preferences, it plays no role in the representation of ex ante preference. Its role is exclusively to represent updating. The divorce from prior beliefs of the likelihoods used for updating does not contradict WDC: prior beliefs about signals underlie choice, but since in WDC signals are assumed not to be payoff relevant, consistency across time does not require that they play a role when processing signals.

Turn to the connection with updating in a Bayesian model. Given a likelihood function L and prior μ as in the theorem, define $\overline{L} \in \Delta(\Omega)$ by

$$\overline{L}(\cdot) = \int L(\cdot \mid V) d\mu(V). \tag{6.4}$$

Note that then the one-step-ahead conditional of $\overline{L}$ at stage n is $\overline{L}_n$ defined by (6.2). It follows that the identical process $\{\mu_n\}$ arises in an expected utility model where $\overline{L}(\cdot)$ is the Bayesian prior.[19] This is not to say that our model is observationally equivalent to the corresponding Bayesian model - they involve the identical process of posteriors but the two models of choice are distinct. For example, only in the shadow Bayesian model do ex ante and conditional preferences satisfy the Independence axiom; in our model preferences at node n are represented by the mutliple-priors utility function $\int V d\mu_n(V)$. The existence of a shadow Bayesian model is an advantage in terms of tractability, since it permits application of results from the Bayesian literature about the dynamics of posteriors.

The emergence of *additive* likelihood functions in spite of the presence of ambiguity should by now not be surprising. At the functional form level, it is a consequence of

[19] Without further assumptions, $\overline{L}$ need not be exchangeable. Thus the shadow Bayesian model is not de Finetti's in general.

preferences being represented by additive measures μ_n. The latter, in turn, emerges as a consequence of Orthogonal Independence. We pointed out when discussing OI that it rules out (in the coin-tossing example) ambiguity about the physical bias of the coin - hedging gains arise only from the poorly understood idiosyncratic factors that affect experiments and render them nonidentical.

Finally, consider briefly uniqueness properties. Define the process $\{w_n\}$ by

$$w_n(s_n; V) = \frac{L_n(s_n \mid V)}{\overline{L}_n(s_n)} = \frac{d\mu_n}{d\mu_{n-1}}. \tag{6.5}$$

Refer to $w_n(s_n, V)$ as the *weight of evidence* for V provided by s_n (and the suppressed s_1^{n-1}). Then the weight of evidence process is unique (up to nullity), because $\{\mu_n\}$ is unique and hence so are the Radon-Nikodym densities $\frac{d\mu_n}{d\mu_{n-1}}$.

On the other hand, the likelihood function L is typically *not unique*. Suppose, for example, that signals are perceived to be uninformative, so that $\mu_n = \mu$ for all n. Then any specification with $L_n(\cdot; V) = \overline{L}_n(\cdot)$, where the latter measures are arbitrary, satisfies (6.1). On the other hand, if for each history s_1^{n-1}, the conditional utility functions $U_n(\cdot \mid s_n)$, $s_n \in S_n$, are linearly independent, then it follows immediately from (6.3) that $\{\overline{L}_n(\cdot)\}$ is unique; and thus the conditional likelihoods $L_n(\cdot; V) = w_n(\cdot; V)\,\overline{L}_n(\cdot)$ are also unique for each s_n. Uniqueness of L follows (up to μ-nullity).

We summarize the preceding more formally. First, we add the axiom:[20]

NON-COLLINEARITY: For each n, the collection $\{U_n(\cdot \mid s_1^n) : s_1^n \in S^n\}$ is linearly independent, where each function $U_n(\cdot \mid s_1^n)$ is viewed as a function on $\mathcal{F}_{>n}$.

Corollary 6.2. *Let L' and L be two likelihood functions that satisfy the conditions in Theorem 6.1. Then, for every n,*

$$w'_n(\cdot; v) = w_n(\cdot; v) \quad \mu_{n-1}\text{-a.s.}$$

where the weights processes $\{w'_n\}$ and $\{w_n\}$ are defined as in (6.5). Moreover, if Non-Collinearity is satisfied, then $L'(\cdot \mid V) = L(\cdot \mid V)$ μ-a.s.

6.2. The Dynamics of Beliefs

The preceding section defines a rich framework for modeling updating - there is room for more structure to be imposed on updating via additional axioms on preferences that restrict the likelihood function L provided by Theorem 6.1. Rather than pursuing further axiomatizations here, we turn instead to illustrating what the model can deliver.

Two properties are immediate and apply at a very general level: (i) Ambiguity is in general not monotonic along a sample. Posterior probabilities $\mu_n(V)$ are not monotonic under Bayesian updating. Thus, for example, if μ has two points of support V' and V, and if $V'(\cdot) \leq V(\cdot)$, (the set of priors for V' includes that for V), then the set of priors

[20] Recall that utilities are "probability equivalents", and thus it is legitimate to use $U_n(\cdot \mid s_1^n)$ in an axiom for conditional preference.

corresponding to U_n decreases with n (in the sense of set inclusion) if $\mu_n(V)$ increases but increases in size if $\mu_n(V)$ decreases. (ii) Ambiguity need not vanish asymptotically (this is illustrated and discussed further below).[21]

To say more, and by way of illustration, we adopt a number of specializations. First, we assume that μ_0, representing ex ante beliefs, has support on Walley-Fine IID utility functions, that is, ex ante utility is given, as in (5.5), by

$$U_0(f) = \int \left(\inf_{\mathcal{L}^\infty} Pf\right) d\mu_0(\mathcal{L}).$$

For concreteness, and to aid interpretation, we consider coin-tossing, $S = \{H, T\}$, though considerable generalization is possible. Then, as pointed out following (5.5), each set $\mathcal{L}$ can be identified with a probability interval $\mathcal{I}_\mathcal{L}$ for Heads, and beliefs μ_n are defined over the set of all intervals contained in $[0, 1]$. Above, the likelihood function $L(\cdot \mid V)$ used for updating was conditioned on the IID utility function V. Here, the latter is in one-to-one correspondence with a set $\mathcal{L}$, and hence with a probability interval for Heads. Thus we can write the updating rule (6.1) in the form:

$$\frac{d\mu_n(\mathcal{I}')}{d\mu_n(\mathcal{I})} = \frac{L_n(s_n \mid \mathcal{I}')}{L_n(s_n \mid \mathcal{I})} \frac{d\mu_{n-1}(\mathcal{I}')}{d\mu_{n-1}(\mathcal{I})},$$

for all intervals $\mathcal{I}', \mathcal{I}$. The interpretation is that beliefs about the probability intervals evolve according to the reweighting described by the likelihood ratio $\frac{L_n(s_n|\mathcal{I}')}{L_n(s_n|\mathcal{I})}$. This is just as in Bayesian updating of beliefs about the relevant parameter, which here is a probability interval for Heads. (To remind the reader, the coin is represented by an interval because the physical bias of the coin is only part of the story - tossing technique is thought to be important to a degree corresponding to the length of the probability interval.)

We specialize the likelihood function by assuming that for μ_0-almost every $\mathcal{I}$ (or $\mathcal{L}$):

L1 $L(\cdot \mid \mathcal{I})$ is exchangeable. Then, by the de Finetti Theorem,

$$L(\cdot \mid \mathcal{I}) = \int_{\Delta(\{H,T\})} \ell^\infty(\cdot)\, d\lambda_\mathcal{I}(\ell), \tag{6.6}$$

for a unique probability measure $\lambda_\mathcal{I}$ on $\Delta(\{H, T\})$.

L2 $\lambda_\mathcal{I}$ has support equal to $\mathcal{I}$. (Here and below we identify $\lambda_\mathcal{I}$ also with a measure on $[0, 1]$ in the obvious way.)

L1 asserts that even conditional on the probability interval $\mathcal{I}$, there is still uncertainty, represented by $\lambda_\mathcal{I}$, about which i.i.d. law describes experiments. Acemoglu *et al* [1] study updating in a completely Bayesian model where likelihoods are specified as in (6.6) in order to capture situations where, for agents trying to learn about $\mathcal{I}$, signals are difficult to interpret. In their case, $\mathcal{I}$ is an abstract parameter rather than a probability interval. Here signals are difficult to interpret exactly because experiments are not identical.

[21] Though our model does not permit infinite samples, asymptotic results can be interpreted as approximations for large finite samples.

L2 asserts that when drawing inferences from a signal about a particular interval $\mathcal{I}$, that is, when conditioning on $\mathcal{I}$, the i.i.d. laws taken into account are precisely those for which the probability of Heads lies in the interval. (An immediate implication is that $L(\cdot \mid \mathcal{I}) \in clh(\mathcal{L}^{\infty})$.)

If $\mathcal{I}$ is the degenerate interval at $p \in [0,1]$, then L1 implies that $L(\cdot \mid p)$ is the i.i.d. measure with probability of Heads equal to p. If the preceding obtains for every $\mathcal{I}$ in the support of μ_0, de Finetti's model, including Bayesian updating, is obtained.

We can now state a counterpart for our framework of the Savage result that data eventually swamp the prior.

Proposition 6.3. *Suppose that the likelihood function L satisfies L1 and L2, and that μ_0 has finite support. (i) Suppose further that for any $\mathcal{I}' \neq \mathcal{I}$ in the support, $\mathcal{I}'$ and $\mathcal{I}$ are disjoint. Then, for every $\mathcal{I}$ with $\mu_0(\mathcal{I}) > 0$,*

$$\mu_n(\mathcal{I}) \to 1 \quad L(\cdot \mid \mathcal{I})\text{-a.s.}$$

(ii) Let μ_0 have support $\{\mathcal{I}, p\}$, where $p(H) \in \mathcal{I}$ is permitted. If p is not an atom of $\lambda_{\mathcal{I}}$, that is, if $\lambda_{\mathcal{I}}(p) = 0$, then

$$\mu_n(p) \to 1 \quad p^{\infty}\text{-a.s.}$$

Part (i) is the indicated counterpart. The assumption of disjoint intervals is an intuitive identification assumption. The set G of samples along which $\mu_n(\mathcal{I})$ converges to 1 satisfies $L(G \mid \mathcal{I}) = 1$, and hence also

$$\ell^{\infty}(G) = 1 \quad \lambda_{\mathcal{I}}\text{-a.s.}$$

Since $\lambda_{\mathcal{I}}$ has full support (L2), this clarifies the sense in which G is a large set.

Note that even given certainty about $\mathcal{I}$, in general there remains ambiguity when predicting future experiments and ranking bets over their outcomes. For example, an individual could become certain about the physical bias of the coin, but in general remain ambiguous about the outcomes of future experiments because of her limited understanding of the effects of tossing technique, particularly her view that these are unrelated across tosses. On the other hand, if the truth is that experiments are i.i.d. with probability of Heads equal to p, if the truth has positive subjective probability ex ante (p is in the support of μ_0), and if the identification condition is satisfied, then the individual asymptotically becomes certain of the true law with probability 1 according to the truth, and there is no ambiguity remaining (she uses the i.i.d. measure corresponding to p to predict future outcomes).

Part (ii) is an illustrative result for the case when intervals may overlap. Here there is convergence to the truth, though the prior attaches positive probability to $\mathcal{I}$, and hence to experiments differing. The overall message is that whether or not ambiguity persists asymptotically depends (on the sample and) on the prior view of experiments. If the individual is certain that each new coin-toss is influenced by a different and hard-to-understand technique, then, even after learning the coin's bias, it is rational to take this limited understanding into account for further prediction and choice. On the other hand, the model does not force ambiguity to persist in all circumstances.

A final example exploits the fact that the "parameters" $\mathcal{I}$ being learned about are probability intervals. Let μ_0 have support $\{\mathcal{I}', \mathcal{I}\}$, where

$$\mathcal{I}' = [p - \delta', p + \delta'],\ \mathcal{I} = [p - \delta, p + \delta] \text{ and } \delta' > \delta > 0.$$

Thus the intervals have a common midpoint but differ in length. Accordingly, we interpret the individual as entertaining two hypotheses that differ *only* in how similar experiments are seen to be; obviously, they are more similar according to $\mathcal{I}$. We ask how the posterior probability $\mu_n(\mathcal{I})$ behaves in large samples.

Specialize L1-L2 by assuming further that $\lambda_{\mathcal{I}'}$ and $\lambda_{\mathcal{I}}$ are uniform on their respective intervals. Though we do this for concreteness, the uniform distribution seems natural. It delivers the following result for the limiting probability of $\mathcal{I}$:[22] Denote by $\Omega_{\mathcal{I}}$ the set of samples ω for which $\lim \Psi_n(\omega) \in \mathcal{I}$. Then, for every ω in $\Omega_{\mathcal{I}}$,

$$\mu_\infty(\mathcal{I}) = \frac{1}{1 + \frac{\mu_0(\mathcal{I}')}{\mu_0(\mathcal{I})} \frac{\delta}{\delta'}}. \tag{6.7}$$

Note that, by (D.1) and the full support property L2, the set of samples $\Omega_{\mathcal{I}}$ has positive probability according to both $L(\cdot \mid \mathcal{I}')$ and $L(\cdot \mid \mathcal{I})$.

For samples in $\Omega_{\mathcal{I}}$, the limiting empirical frequency of Heads is consistent with *both* $\mathcal{I}$ and $\mathcal{I}'$. This identification problem leads to the result that $0 < \mu_\infty(\mathcal{I}) < 1$ - neither hypothesis is dismissed entirely along such samples, even in the limit. This is an instance of the identification problem studied by Acemoglu *et al* [1]. In spite of differences between the two models, some of their other results also translate into our setting. In particular, one could use concern about nonidentical experiments to justify asymptotic disagreement between individuals.

Another noteworthy implication of (6.7) is that $\mu_\infty(\mathcal{I}) > \mu_0(\mathcal{I})$, that is, any sample that is consistent with both hypotheses leads eventually to a shift in probability mass towards the "more precise" hypothesis. Given a sample, the difficulty in making inferences about future experiments is that they are not seen to be identical. Here experiments may differ according to both $\mathcal{I}'$ and $\mathcal{I}$, but they differ more according to $\mathcal{I}'$. Thus the sample provides less information about future experiments under $\mathcal{I}'$ than under $\mathcal{I}$. This leads to a shift in weight towards $\mathcal{I}$.

7. REGULARITY OR MONOTONE CONTINUITY?

Return to the static or one-shot choice setting. With the convention that outcomes are measured in utils, the multiple-priors model is usually written in the form

$$U(f) = \min_{P \in C} \int_\Omega f dP = \min_{P \in C} Pf, \quad f \in \mathcal{F}, \tag{7.1}$$

[22] The claim (6.7) to follow is adapted from Acemoglu *et al* [1, Lemma 1]. The latter implies also that for the lack of asymptotic learning, it would be enough for $\lambda_{\mathcal{I}'}$ and $\lambda_{\mathcal{I}}$ to have positive and continuous Lebesgue densities on their intervals.

where $C \subset ba^1_+(\Omega)$ is a convex and weak*-compact set of *finitely* additive probability measures on (Ω, Σ).[23] Gilboa and Schmeidler [19] prove that this model is characterized by a simple set of axioms.

At the functional form level, our version of multiple-priors (2.1) is evidently the special case where the Gilboa-Schmeidler set of priors C is the weak*-closure of a convex and weak-convergence closed set $\mathcal{P}$ of *countably* additive priors. Our main objective in this section is to describe the behavioral meaning of this specialization.

The rationale for the specialization is straightforward - just as countable additivity is assumed widely in the central theorems of probability theory, including in the de Finetti Theorem that concerns us here, we specialize multiple-priors utility to provide a counterpart of countable additivity.[24] Since Chateauneuf *et al* [7] put forth a more restrictive way to express "countable additivity" for a set of priors, we examine it in some detail and argue that, though it is in some sense simpler, it is unduly restrictive particularly for a setting with repeated experiments.

7.1. Regularity

The added axiom that we impose on preference, or utility, is Regularity, a property first studied in Epstein and Wang [17]. Roughly, it extends to preferences the well-known property of regularity of probability measures.[25] A connection to countable additivity is that any measure on a compact metric space is countably additive if and only if it is regular [14, p. 138]. Thus it is not easy to distinguish between these properties within the space of measures. However, as will become evident, they lead to substantially different notions more generally, and we argue that there is a distinct advantage to using regularity to define the ambient technical framework.

The set of all $[0,1]$-valued acts on Ω is $\mathcal{F}$. Denote by $\mathcal{F}^u$ the set of all upper semicontinuous (usc) and simple (finite-ranged) acts, and by $\mathcal{F}^\ell$ the set of lower semicontinuous (lsc) and simple acts. As shown above, under suitable conditions there is a unique probability-equivalent utility function U, defined in (2.3), that represents preference. Thus we can state the sought-after condition in terms of that utility function.[26]

REGULARITY: A utility function $U : \mathcal{F} \to [0,1]$, and the corresponding preference order, are regular if both of the following conditions are satisfied:

Inner Regularity $U(h) = \sup\{U(g) : g \leq h, g \in \mathcal{F}^u\}$, $\forall h \in \mathcal{F}^\ell$; and

[23] The following additional notation is needed here. For any compact metric space X, $ba(X)$ and $ca(X)$ denote the spaces of finite variation set functions on the Borel σ-algebra that are finitely additive (charges) and countably additive respectively; $ba^1_+(X)$ and $ca^1_+(X)$ are the corresponding subsets of positive and normalized measures. The notation $ca^1_+(X)$, in place of $\Delta(X)$, is useful when it is important to draw a distinction between finitely and countably additive probability measures. Unless otherwise specified, the weak-convergence topology is used for $ca^1_+(X)$. By the weak* topology on $ba(\Omega)$, we mean the topology induced by bounded measurable functions.

[24] See Regazzini *et al* [32] and Dubins [12, 13] for approaches assuming only finite additivity.

[25] The reader is referred to [17] for detailed discussion of regularity of preferences and the formal relationship to regular probability measures and also regular capacities.

[26] We state Regularity for any utility function U. As shown in [17], the axiom is readily expressed explicitly in terms of preference for a large class of preferences.

Outer Regularity $U(f) = \inf\{U(h) : h \geq f, h \in \mathcal{F}^{\ell}\}, \forall f \in \mathcal{F}$.

To see the parallel with the notion of regularity for a measure, think of the special case of acts that are indicator functions, and note that the indicator $\mathbf{1}_A$ is simple and usc (lsc) if A is closed (open). This parallel inspired the closely related, but distinct, definition of regularity of preference in [17]. The relation is that U is regular in the above sense if and only if its conjugate U^*,

$$U^*(f) = 1 - U(1-f), \ \ f \in \mathcal{F},$$

is regular in the sense of Epstein and Wang. For another perspective on the difference between the two definitions of regularity, observe that the Epstein-Wang notion requires that

$$U(f) = \sup\{U(g) : g \leq f, g \in \mathcal{F}^u\}, \forall f \in \mathcal{F},$$

that is, the utility of arbitrary acts can be approximated from below (by simple usc acts). In contrast, Outer Regularity above postulates that the utility of arbitrary acts can be approximated from above (by simple lsc acts). Approximation from above seems more intuitive given the conservatism inherent in aversion to ambiguity or to limited evidence. Since any probability measure coincides with its conjugate ($P(A) = 1 - P(\Omega \backslash A)$), the two notions of regularity coincide in the SEU case, where $U(f) = Pf$ for a fixed P, with the usual notion of regularity of the measure P.[27]

We can now state the main result of this subsection, which shows that Regularity characterizes our specialization of the Gilboa-Schmeidler model.[28]

Theorem 7.1. *Let U be a multiple-priors utility as in (7.1). Then U satisfies Regularity if and only if it can be expressed in the form (2.1) for some $\mathcal{P} \subset ca^1_+(\Omega)$ that is convex and (weak-convergence) compact. Moreover, the set $\mathcal{P}$ is unique.*

An important implication of Regularity is that utility is completely determined by its values on finitely-based acts. Much as the Kolmogorov Extension Theorem tells us that a probability measure on $\Omega = S^\infty$, (which is necessarily regular given that Ω is metric), is completely determined by its values on finite cylinders, a generalized extension theorem proven in [17, Theorem D.2] implies that a regular utility is uniquely determined by its values on $\mathcal{F}_{fin}$.[29] This feature of our model of multiple-priors utility was behind the scenes of a number of results stated above.

For example, consider the proof of Proposition 3.1. Symmetry implies that preference over $\mathcal{F}_{fin}$ is represented both by $\mathcal{P}$ and by $\{\pi P : \pi \in \Pi, P \in \mathcal{P}\}$. Hence they represent

[27] More precisely, it follows from [17, Theorem 4.1] that: an SEU preference with prior P is regular in the sense of Epstein-Wang if and only if it is regular in the sense of this paper if and only if P is a regular measure.

[28] See Appendix A for a proof; it relies, for one direction, on a result by Chen [8]. We remind the reader that since the probability-equivalent utility U corresponds uniquely to preference, the theorem could be restated in terms of the latter. Finally, see Philippe *et al* [31, Proposition 1] for a related result dealing with lower envelopes of sets of priors rather than with preferences over acts.

[29] The different meaning of "regularity," explained above, does not affect the validity of the Kolmogorov-style theorem.

the same preference over $\mathcal{F}$ by [17, Theorem D.2]. Therefore, they must be identical by the uniqueness of the representing set of priors (Theorem 7.1).

As a second example, consider a gap in the proof that (i) implies (ii) in Theorem 3.2. Let $\widehat{U}(f) = U(\alpha f + (1-\alpha)\pi f)$. Strong Exchangeability implies $\widehat{U} = U$ on $\mathcal{F}_{fin}$. Lemma A.1 shows that $\widehat{U}$ satisfies Regularity. Therefore, $\widehat{U}(f) = U(f)$ for all $f \in \mathcal{F}$ by Epstein and Wang [17, Theorem D.2].

7.2. An Alternative: Monotone Continuity

An alternative way to express "countable additivity" for a set of priors, put forth by Chateauneuf *et al* [7], is to assume that $\mathcal{P}$ itself is weak*-compact, and hence that C (equals $\mathcal{P}$ and) consists exclusively of countably additive measures. What could be a more natural way to formulate the counterpart of countable additivity of single measures?

It may seem plausible also at the more meaningful behavioral level. Chateauneuf *et al* show that weak*-compactness of $\mathcal{P}$ is characterized behaviorally by Monotone Continuity:

MONOTONE CONTINUITY: Given $f \succ g$, outcome x, and a sequence $\{A_n\}$ in Σ, with $A_n \searrow \varnothing$, then, there exists N such that $(x, A_N; f(\cdot), \Omega \backslash A_N) \succ g$ and $f \succ (x, A_N; g(\cdot), \Omega \backslash A_N)$.

As the cited authors point out, this axiom is used by Arrow [3] to characterize countable additivity of the Savage prior. Moreover, Monotone Continuity is arguably simpler than Regularity.[30]

On the other hand, Monotone Continuity is *stronger* than Regularity, (weak*-compactness implies weak-convergence compactness for any set of priors), and we argue that *the difference is significant.*

For example, Monotone Continuity implies that[31]

$$U(B_n) \nearrow U(B) \text{ for } \textbf{all} \text{ sequences } B_n \nearrow B. \tag{7.2}$$

In contrast, Regularity (via Inner Regularity) imposes only that $U(B)$ can be approximated from below by $U(B_n)$ by **some** sequence B_n that increases to B. (In fact, this is required only if B is open, which notably excludes B being a tail event; on the other hand, the approximating sets must be compact.) To see that the difference between "for all" and "for some" can be significant, consider the following coin-tossing example. The coin is known to be unbiased, but there remains uncertainty surrounding the tossing techniques of different people. You believe that every person imparts an (idiosyncratic) effective bias lying in $\{0, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...\}$, but are completely ignorant within this set. A model that captures this perception is the IID utility function corresponding to $\mathcal{L}^\infty$, where

$$\mathcal{L} = \{\delta_0\} \cup \{\delta_{1/n} : n > 1\} \subset \Delta(\{H, T\}).$$

[30] Monotone Continuity is definitely easier to state, but it is not clear that its meaning is easier to grasp. For example, the Borel σ-algebra includes many complicated events that are difficult even to describe. Hence the scope of a condition that applies to all (measurable) acts is hard to understand. The surprising Theorem 7.2 below illustrates this point.

[31] Let $B_n \nearrow B$. Define $A_n = B \backslash B_n \searrow \varnothing$, $f = 1_B$ and $f_n = 1_{B_n} = (0, A_n; f, \Omega \backslash A_n)$. Then Monotone Continuity implies that, for every $\epsilon > 0$, there exists N such that $V(B_N) = V(f_N) > (1-\epsilon) V(B)$.

Let

$$B_n = \{\omega : \lim_k \Psi_k(\omega) = 0, \text{ or } \limsup_k \Psi_k(\omega) \geq \frac{1}{n}\},$$

where $\Psi_k(\omega)$ denotes the empirical frequency of Heads in the first k tosses along the sample ω. Observe that $B_n \nearrow \Omega$. However, $U(B_n) = 0$ for every n: the set of priors $\mathcal{L}^\infty$ includes an i.i.d. measure Q where Heads has probability λ in $\left(0, \frac{1}{n}\right)$, and thus according to which the empirical frequency of Heads converges with certainty to λ. Therefore,

$$U(B_n) = \inf_{P \in \mathcal{L}^\infty} P(B_n) \leq Q(B_n) = 0.$$

You would not be willing to bet on B_n because, no matter how large is n, the worst-case scenario is that many people impart a bias smaller than $\frac{1}{n}$, and this would lead to a sample path not in B_n. On the other hand, of course, $U(\Omega) = 1$, violating Monotone Continuity. In contrast, there is no contradiction to Regularity, since Ω can be approximated from below by some sequence $\{K_n\}$ of compact sets - $K_n = \Omega$ for all n works trivially.

More generally, while Regularity is consistent with the Walley-Fine IID utility function (5.3), that model is excluded if Monotone Continuity is assumed, because $\mathcal{L}^\infty$ is not weak*-compact unless $\mathcal{L}$ is a singleton. Here is a proof: For simplicity, consider $S = \{H, T\}$. Let $\ell_0, \ell_1 \in \mathcal{L}$ and $\ell_0 \neq \ell_1$. For any $r \in [0,1]$, we can find $\{i_t\}_{t=1}^\infty \in \{0,1\}^\infty$ such that $\frac{1}{N}\sum i_t$ converges to r. Then, by Hall and Heyde [20, Theorem 2.19], the measure $\otimes_t \ell_{i_t} \in \mathcal{L}^\infty$ assigns 1 to the event A_r, where A_r is the set that the limiting empirical frequency of Head is $(1-r)\ell_0(H) + r\ell_1(H)$. If $\mathcal{L}^\infty$ were weak*-compact, there would be $Q \in \Delta(S^\infty)$ such that $Q(A) = 0$ implies $P(A) = 0$ for all $P \in \mathcal{L}^\infty$, by [7, Lemma 3]. Thus, $Q(A_r) > 0$ for all $r \in [0,1]$, which cannot be true.

The Walley-Fine utility function (5.3) is only one example of an IID utility function, that is, a function satisfying Symmetry and the product rule (5.1). Finally, we show that Monotone Continuity excludes *all* IID utility functions, other than expected utility functions. Thus in the setting of infinitely many experiments that are viewed symmetrically, Monotone Continuity excludes modeling the perception that experiments are unrelated in the natural sense of (5.1).

Theorem 7.2. *If V is an IID utility function that satisfies Monotone Continuity, then V is an expected utility function (with an i.i.d. prior).*

The key to the proof of the theorem (found in Appendix C) is to show that Monotone Continuity, plus Symmetry and the stochastic independence condition (5.1), imply that V is 0-1 valued *and* additive on Σ^{tail}. That implies that all measures in $\mathcal{P}$ are 0-1 valued and that they agree on Σ^{tail}. The rest is straightforward.

The restrictiveness of Monotone Continuity is not limited to settings with repeated experiments. For example, let the state space be $[0,1]$ and consider the set of priors $\mathcal{P}$ equal to the weak-convergence closed convex hull of $\{\delta_0\} \cup \{\delta_{1/n} : n > 1\}$. Thus the true state is known to lie in $\{0, \frac{1}{2}, ..., \frac{1}{n}...\}$, but there is complete ignorance within the set. Then Monotone Continuity is violated along the sequence $B_n = \{0\} \cup [\frac{1}{n}, 1] \nearrow [0,1]$, since $U(B_n) = 0 \not\longrightarrow 1 = U(\Omega)$. This reflects an inherent discontinuity arising from ignorance. Again, Inner Regularity is trivially satisfied at Ω.

A. Appendix: Regularity

Proof of Theorem 7.1:
$\Longleftarrow$: Chen [8, Proposition 1] proves that V is regular in the sense of Epstein-Wang, where

$$V(f) = \sup_{P \in \mathcal{P}} Pf = \max_{P \in cl(\mathcal{P})} Pf, \; f \in \mathcal{F}.$$

($cl^*(\mathcal{P})$ denotes the weak*-closure of $\mathcal{P}$ in $ba_+^1(\Omega)$.) Therefore, $U = V^*$, the conjugate of V, satisfies Regularity.

$\Longrightarrow$: The multiple-priors utility function U can be extended in the obvious way to $C(\Omega)$, the set of all continuous real-valued functions on Ω, and the extension is norm-continuous, superadditive, monotone, and $U(1) = 1$. Therefore, it is a support function for a unique compact and convex set $\mathcal{P} \subset ca_+^1(\Omega)$. In particular,

$$U(f) = \min_{\mathcal{P}} Pf, \text{ for every continuous act } f.$$

Let $g \in \mathcal{F}^u$. By Outer Regularity, there exist $h_i \in \mathcal{F}^\ell$ such that

$$h_i \geq g \text{ and } U(h_i) < U(g) + 2^{-i}.$$

Further, there exist continuous acts f_i such that

$$h_i \geq f_i \geq g.$$

(When $h_i = \mathbf{1}_{G_i}$ and $g = \mathbf{1}_K$ are indicator acts, this follows from Urysohn's Lemma. More generally, the assertion follows from a straightforward extension of Urysohn's Lemma for simple acts - see [17, Lemma A.1].) Finally, it is wlog to assume that $f_i \searrow g$ (see Aliprantis and Border [2, Theorem 3.13]). It follows that

$$U(g) = \inf_i U(f_i) = \inf_i \inf_{\mathcal{P}} Pf_i = \inf_{\mathcal{P}} \inf_i Pf_i = \inf_{\mathcal{P}} Pg;$$

the last equality follows because $\inf_i Pf_i = Pg$ for every P by the Monotone Convergence Theorem.

Define

$$\overline{U}(f) = \inf_{\mathcal{P}} Pf = \min_{cl(\mathcal{P})} Pf, \; f \in \mathcal{F}.$$

By the first part of the proof, $\overline{U}$ is regular, while U is regular by assumption. As just shown, the two utility functions agree on $\mathcal{F}^u$. It follows immediately from Regularity that they must agree on all of $\mathcal{F}$. By the uniqueness of the (weak*-compact and convex) set of priors, proven by Gilboa and Schmeidler, C is the weak* closure of $\mathcal{P}$. ∎

The following lemma was used in the proof of Theorem 3.2.

Lemma A.1. *If U is regular, then so is $\widehat{U}$, where $\widehat{U}(f) = U(\alpha f + (1-\alpha)\pi f)$, $f \in \mathcal{F}$, for any fixed α and π.*

Proof. Inner Regularity: Take $h \in \mathcal{F}^{\ell}$. It is clear that $U(h) \geq \sup\{U(g) : g \leq h, g \in \mathcal{F}^{u}\}$. Next show equality. By [2, Theorem 3.13], we can take continuous f_n such that $f_n(\omega) \nearrow h(\omega)$ for each $\omega \in \Omega$. By [17, Lemma D.3], there exist finitely-based h'_n such that $f_n \leq h'_n \leq h$. Thus,

$$\begin{aligned}
\lim_n U\left(\alpha h'_n + (1-\alpha)\,\pi h'_n\right) &= \lim_n \min_{P\in\mathcal{P}} \int \left(\alpha h'_n + (1-\alpha)\,\pi h'_n\right) dP \\
&= \min_{P\in\mathcal{P}} \lim_n \int \left(\alpha h'_n + (1-\alpha)\,\pi h'_n\right) dP \\
&= \min_{P\in\mathcal{P}} \int \left(\alpha h + (1-\alpha)\,\pi h\right) dP \\
&= U\left(\alpha h + (1-\alpha)\,\pi h\right).
\end{aligned}$$

The second equality follows from Terkelsen's minimax theorem [36, Corollary, p.407] and the third by the Monotone Convergence Theorem for each P.

Outer Regularity: Note that

$$\begin{aligned}
U\left(\alpha f + (1-\alpha)\,\pi f\right) &= \inf_{P\in\mathcal{P}} \int \left(\alpha f + (1-\alpha)\,\pi f\right) dP \\
&= \inf_{P\in\mathcal{P}} \inf_{f\leq h\in\mathcal{F}^{\ell}} \int \left(\alpha h + (1-\alpha)\,\pi h\right) dP \\
&= \inf_{f\leq h\in\mathcal{F}^{\ell}} \inf_{P\in\mathcal{P}} \int \left(\alpha h + (1-\alpha)\,\pi h\right) dP \\
&= \inf_{f\leq h\in\mathcal{F}^{\ell}} U\left(\alpha h + (1-\alpha)\,\pi h\right).
\end{aligned}$$

The second equality follows because $f \longmapsto \int \left(\alpha f + (1-\alpha)\,\pi f\right) dP$ satisfies Regularity. ■

B. Appendix: Proof of Theorem 5.2

After proving necessity, the bulk of the proof concerns sufficiency of the axioms. Here we adapt the Hewitt-Savage [23] proof strategy for the de Finetti theorem to our setting. In broad terms, it amounts to showing that the set $\mathcal{U}^*$ of multiple-priors utility functions satisfying Symmetry, OI and Super-Convexity is compact and convex, and then using the Choquet Theorem (Phelps [30, p.14]) to express any such utility function as an integral over extreme points of $\mathcal{U}^*$. The proof of uniqueness concludes.

B.1. Necessity

Show first that the integral $\int_{\mathcal{V}} V(f)\, d\mu(V)$ is well-defined for all f in $\mathcal{F}$. Denote by

$$\mathcal{Q}^{IID} = \{\mathcal{P} \in \mathcal{K}^c(\Delta(\Omega)) : U_{\mathcal{P}} \in \mathcal{V}\}$$

the set of all IID sets of priors. We show below that $\mathcal{V}$, and hence also $\mathcal{Q}^{IID}$, are compact, hence Borel measurable. Since μ is well-defined on $\overline{\Sigma}$, the universal completion of Σ, it suffices to show that the function $V \longmapsto V(f)$ is universally measurable. This is true if every set of the form

$$\left\{\mathcal{P} \in \mathcal{Q}^{IID} : \exists P \in \mathcal{P}, Pf < c\right\} =$$
$$proj_{\Delta(\Omega)}\left(\left\{(\mathcal{P}, P) \in \mathcal{K}^c(\Delta(\Omega)) \times \Delta(\Omega) : \mathcal{P} \in \mathcal{Q}^{IID}, P \in \mathcal{P}, Pf < c\right\}\right)$$

lies in $\overline{\Sigma}$. But the set being projected is Borel-measurable (in the product σ-algebra). Therefore, the projection is universally measurable by the Lusin-Choquet-Meyer Theorem [24, Theorem A.1.8, p. 457].

Define U by (5.4), that is,

$$U(f) = \int_{\mathcal{V}} V(f)\, d\mu(V), \text{ for all } f \text{ in } \mathcal{F}.$$

The Gilboa-Schmeidler axioms are clearly satisfied. OI and Super-Convexity can be proven as in the proof of Lemma 5.1. We need to show that U is a (regular) multiple-priors utility function. To do so, we establish a suitable set of priors for U.

Since $\mathcal{U}$ is homeomorphic to $\mathcal{K}^c(\Delta(\Omega))$, $\mu \in \Delta(\mathcal{V})$ can be viewed as a measure on $\mathcal{K}^c(\Delta(\Omega))$. Thus, we can write

$$U(f) = \int U_{\mathcal{Q}}(f)\, d\mu(\mathcal{Q}). \tag{B.1}$$

We define the Aumann integral $\int \mathcal{Q} d\mu(\mathcal{Q})$ as follows: For a measurable $\phi : \mathcal{K}^c(\Delta(\Omega)) \to \Delta(\Omega)$, define $\int \phi(\mathcal{Q})\, d\mu(\mathcal{Q}) = \int \phi d\mu \in \Delta(\Omega)$, by[32]

$$\left(\int \phi(\mathcal{Q})\, d\mu(\mathcal{Q})\right)(A) = \int \phi(\mathcal{Q})(A)\, d\mu(\mathcal{Q}) \text{ for all } A \in \Sigma.$$

Let Φ be the identity function from $\mathcal{K}^c(\Delta(\Omega))$ to $\mathcal{K}^c(\Delta(\Omega))$ and $\operatorname{Sel}\Phi$ the set of all measurable selections from Φ, that is, $\phi \in \Phi$ iff ϕ is a measurable function from $\mathcal{K}^c(\Delta(\Omega))$ to $\Delta(\Omega)$ satisfying $\phi(\mathcal{Q}) \in \mathcal{Q}$. Then

$$\int \mathcal{Q} d\mu(\mathcal{Q}) \equiv \left\{\int \phi d\mu : \phi \in \operatorname{Sel}\Phi\right\}.$$

Below we use the next lemma, which can be proven by a standard argument using the Lebesgue Dominated Convergence Theorem.

Lemma B.1. *Let $\phi : \mathcal{K}^c(\Delta(\Omega)) \to \Delta(\Omega)$ be measurable and $P = \int \phi d\mu$. Then, for any $f \in \mathcal{F}$,*

$$\int_{\mathcal{K}^c(\Delta(\Omega))} \left[\int_{\Omega} f d\phi(\mathcal{Q})\right] d\mu(\mathcal{Q}) = \int f dP.$$

[32] The right side is well-defined because $\mathcal{Q} \longmapsto \mathcal{Q}(A)$ is measurable by [2, Lemma 15.16]. It is easy to see that one obtains a countably additive measure.

Lemma B.2. *Let $\mu \in \Delta(\mathcal{K}^c(\Delta(\Omega)))$. Then, for all $f \in \mathcal{F}$,*

$$\inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP = \int_{\mathcal{K}^c(\Delta(\Omega))} \left(\inf_{P \in \mathcal{Q}} \int f dP \right) d\mu(\mathcal{Q}).$$

$$= \int_{\mathcal{K}^c(\Delta(\Omega))} U_{\mathcal{Q}}(f) \, d\mu(\mathcal{Q}) \equiv U(f),$$

where U is defined by (B.1).

Proof. We use a result of Castaldo et al [6, Theorem 3.2], which translated into our setup, states

$$\inf_{\phi \in \mathrm{Sel}\, \Phi} \int_{\Delta(\Omega)} \hat{f}(P) \, d(\mu \circ \phi^{-1})(P) = \int \inf_{P \in \mathcal{Q}} \hat{f}(P) \, d\mu(\mathcal{Q})$$

for any measurable $\hat{f} : \Delta(\Omega) \to \mathbb{R}$. Given $f \in \mathcal{F}$, define $\hat{f}(P) = \int f dP$. Then

$$\begin{aligned}
\inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP &= \inf_{\phi \in \mathrm{Sel}\, \Phi} \int_{\mathcal{K}^c(\Delta(\Omega))} \left[\int_{\Omega} f d\phi(\mathcal{Q}) \right] d\mu(\mathcal{Q}) \\
&= \inf_{\phi \in \mathrm{Sel}\, \Phi} \int_{\mathcal{K}^c(\Delta(\Omega))} \hat{f}(\phi(\mathcal{Q})) \, d\mu(\mathcal{Q}) \\
&= \inf_{\phi \in \mathrm{Sel}\, \Phi} \int_{\Delta(\Omega)} \hat{f}(P) \, d(\mu \circ \phi^{-1})(P) \\
&= \int \inf_{P \in \mathcal{Q}} \hat{f}(P) \, d\mu(\mathcal{Q}) = \int_{\mathcal{K}^c(\Delta(\Omega))} \left[\inf_{P \in \mathcal{Q}} \int f dP \right] d\mu(\mathcal{Q}),
\end{aligned}$$

where the third equality follows by the Change of Variable Theorem [2, Theorem 13.46], and the fourth by the result cited above. ■

Lemma B.3. *Let $\overline{\mu}$ denote the weak-convergence closure of $\int \mathcal{Q} d\mu(\mathcal{Q})$. Then:*
(i) $\overline{\mu} \in \mathcal{K}^c(\Delta(\Omega))$; and
(ii) $\int U_{\mathcal{Q}}(f) \, d\mu(\mathcal{Q}) = \inf_{P \in \overline{\mu}} \int f dP$ for all $f \in \mathcal{F}$.
In other words, the utility function U defined as in (5.4), or equivalently, as in (B.1), is a regular multiple-priors utility function with set of priors $\overline{\mu}$.

Proof. (i) Convexity is clear because if ϕ, ϕ' are measurable selections, then so is any convex combination. The set $\overline{\mu}$ is closed by construction.
(ii) Define

$$\overline{U}(f) = \inf_{P \in \overline{\mu}} \int f dP, \; f \in \mathcal{F}.$$

By (i) and Theorem 7.1, $\overline{U}$ is regular.
Step 1: By the preceding lemma,

$$U(f) \equiv \int U_{\mathcal{Q}}(f) \, d\mu(\mathcal{Q}) = \inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP, \text{ for every } f \in \mathcal{F}.$$

Step 2:

$$\overline{U}(f) = \inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP, \text{ for every lsc } f.$$

It is enough to show that $\min_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP$ exists for every lsc f. By the preceding lemma,

$$\inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP = \int \left(\min_{P \in \mathcal{Q}} \int f dP \right) d\mu(\mathcal{Q}).$$

By the Measurable Maximum Theorem [2, Theorem 18.19], there is a measurable selection ϕ, $\phi(\mathcal{Q}) \in \text{argmin}_{P \in \mathcal{Q}} \int f dP$ for each $\mathcal{Q}$. Then $\overline{P} = \int \phi d\mu \in \int \mathcal{Q} d\mu(\mathcal{Q})$, and

$$\int \left(\min_{P \in \mathcal{Q}} \int f dP \right) d\mu(\mathcal{Q}) = \int f d\overline{P}.$$

Step 3:

$$\overline{U}(f) = \inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP, \text{ for every } f \in \mathcal{F}.$$

Argue as follows:

$$\int \mathcal{Q} d\mu(\mathcal{Q}) \subset \overline{\mu} \Longrightarrow \inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP \geq \overline{U}(f).$$

Next prove the reverse inequality. Since $\overline{U}$ is regular, by Outer Regularity, given any f and ϵ, there exists a simple lsc h such that

$$h \geq f \text{ and } \overline{U}(h) < \overline{U}(f) + \epsilon.$$

But then Step 2 and $h \geq f \Longrightarrow$

$$\inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP \leq \inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int h dP = \overline{U}(h) < \overline{U}(f) + \epsilon,$$

which proves the desired inequality

$$\inf_{P \in \int \mathcal{Q} d\mu(\mathcal{Q})} \int f dP \leq \overline{U}(f).$$

Combine Steps 1 and 3 to complete the proof. ■

B.2. Sufficiency: The Hewitt-Savage Strategy Adapted

We turn to the sufficiency part of the theorem. Assume Symmetry, OI and Super-Convexity.

We exploit heavily the homeomorphism between $\mathcal{K}^c(\Delta(\Omega))$, the space of sets of priors, and $\mathcal{U} = \{U_{\mathcal{P}} : \mathcal{P} \in \mathcal{K}^c(\Delta(\Omega))\}$, the space of (regular) multiple-priors utility functions. We pass freely between them. Recall also that $\mathcal{K}^c(\Delta(\Omega))$, and hence also $\mathcal{U}$, are compact metric.

The following preliminary results are straightforward.

Lemma B.4. *$\mathcal{P} \longmapsto \min_{P \in \mathcal{P}} Pf$ is continuous for any continuous act f.*

Proof. This is implied by the Maximum Theorem [2, Theorem 17.31]. ■

Define

$$\mathcal{U}^* = \{U \in \mathcal{U} : U \text{ satisfies Symmetry, OI and Super-Convexity}\}.$$

Lemma B.5. *$\mathcal{U}^*$ is compact and convex, and $\mathcal{V}$, the subset of IID utility functions, is compact.*

Proof. As noted, $\mathcal{U}$ is compact. The further defining properties of $\mathcal{U}^*$ and $\mathcal{V}$ deal with finitely-based, and hence continuous, acts only. Therefore, the preceding lemma implies that each set is closed. Convexity of $\mathcal{U}^*$ is obvious. ■

The following lemma is the key to identifying the extreme points of $\mathcal{U}^*$. Much of the next subsection is concerned with proving the lemma. We continue here assuming the lemma is true.

Lemma B.6. *For any $U \in \mathcal{U}^*$ and $f^* \in \mathcal{F}_{\{1,\ldots,m\}}$ satisfying $U(f^*) \in (0,1)$, define the functions U^* and U^{**} by: for all $f \in \mathcal{F}$,*

$$\begin{aligned} U^*(f) &= \frac{U(f^* \cdot \theta^m f)}{U(f^*)}, \text{ and} \\ U^{**}(f) &= \frac{U(\theta^m f) - U(f^* \cdot \theta^m f)}{1 - U(f^*)}. \end{aligned} \tag{B.2}$$

Then $U^*, U^{**} \in \mathcal{U}^*$.

Proposition B.7. *If U is an extreme point of $\mathcal{U}^*$, written $U \in ext(\mathcal{U}^*)$, then $U \in \mathcal{V}$.*[33]

Proof. Let $U \in ext(\mathcal{U}^*)$. It suffices to show that

$$U(f^* \cdot \theta^m f) = U(f^*) U(f), \tag{B.3}$$

for every $f^* \in \mathcal{F}_{\{1,\ldots,m\}}$ and $f \in \mathcal{F}_{fin}$.

Let $\mathcal{P} \in \mathcal{K}^c(\Delta(S^\infty))$ be the set of priors corresponding to U. Consider three cases.

Case 1: $U(f^*) = 0$. Then $\int f^* dP = 0$ for some $P \in \mathcal{P}$. Therefore,

$$f^* \geq 0 \Longrightarrow f^*(\omega) = 0,\ P\text{-}a.s. \Longrightarrow (f^* \cdot \theta^m f)(\omega) = 0,\ P\text{-}a.s.,$$

which implies (B.3).

Case 2: $U(f^*) = 1$. Then $\int f^* dP = 1$ for all $P \in \mathcal{P}$, and, again for all P,

$$f^* \leq 1 \Longrightarrow f^*(\omega) = 1,\ P\text{-}a.s. \Longrightarrow (f^* \cdot \theta^m f)(\omega) = \theta^m f(\omega),\ P\text{-}a.s.$$

[33] We show later that the converse is also true - $ext(\mathcal{U}^*) = \mathcal{V}$ - though we use only the fact that all extreme points lie in $\mathcal{V}$.

Therefore,

$$U(f^* \cdot \theta^m f) = U(\theta^m f) = U(f) = U(f^*)\, U(f),$$

where use has been made of the fact that Symmetry implies "shift invariance":

$$f \sim \theta f \quad \text{for every } f \in \mathcal{F}_{fin}.$$

Case 3: $U(f^*) \in (0,1)$. For every $f \in \mathcal{F}$,

$$U(f) = U(\theta^m f) = U(f^*)\, U^*(f) + (1 - U(f^*))\, U^{**}(f),$$

where U^* and U^{**} are defined in Lemma B.6. Thus

$$U = \alpha U^* + (1-\alpha)\, U^{**}$$

where $\alpha = U(f^*)$. Since U is an extreme point of $\mathcal{U}^*$ and $U^*, U^{**} \in \mathcal{U}^*$, we have $U = U^*$, and hence

$$U(f) = \frac{U(f^* \cdot \theta^m f)}{U(f^*)},$$

especially for $f \in \mathcal{F}_I$ with finite I. This proves (B.3). ■

We wish to apply the Choquet theorem [30, p.14]. For that purpose, note that $\mathcal{U} \subset E \equiv \{\alpha U : \alpha \in \mathbb{R}, U \in \mathcal{U}\}$, where E is a locally convex (vector) space under the topology generated by sets of the form $\{\alpha U : a < \alpha < b, U \in G,\, G \text{ open in } \mathcal{U}\}$. Now take $U \in \mathcal{U}^*$. Then Lemma B.5 and Choquet's theorem imply the existence of a Borel probability measure μ on the set of extreme points of $\mathcal{U}^*$ such that $L(U) = \int L(V)\, d\mu(V)$ for every continuous linear functional L on E. Since $\alpha U \longmapsto \alpha U(f)$ is linear and continuous on E for every continuous f, it follows that

$$U(f) = \int V(f)\, d\mu(V) \tag{B.4}$$

for every continuous f. This, in fact, holds for any $f \in \mathcal{F}$: From the necessity proof, we know that $f \longmapsto \int V(f)\, d\mu(V)$ defines a utility function satisfying Regularity. In addition, U satisfies Regularity by assumption. Finitely-based acts are continuous since S is finite. Thus we can invoke the generalized Kolmogorov extension theorem in Epstein and Wang [17, Theorem D.2] to conclude that (B.4) holds for any $f \in \mathcal{F}$.

This completes the proof of sufficiency in Theorem 5.2, once we have proven Lemma B.6.

B.3. Remaining Arguments Re Extreme Points of $\mathcal{U}^*$

The main objective in this section is to prove Lemma B.6, namely that the two functions U^* and U^{**} defined there lie in $\mathcal{U}^*$.

That $U^* \in \mathcal{U}^*$ is straightforward. First, we show that it is regular.

Lemma B.8. *For any $f^* \in \mathcal{F}_{\{1,\ldots,m\}}$ with $U(f^*) > 0$, the function $U^* : \mathcal{F} \to [0,1]$, defined by*

$$U^*(f) = \frac{U(f^* \cdot \theta^m f)}{U(f^*)}, \quad f \in \mathcal{F},$$

satisfies Regularity.

Proof. Show Outer Regularity. Inner Regularity can be shown in the same way.

View f^* also as a function of $(s_1, ..., s_m) \in S^m$. By Regularity for U, there exist $h_n \in \mathcal{F}^\ell$ such that $h_n \geq f^* \cdot \theta^m f$ and $U(h_n) \searrow U(f^* \cdot \theta^m f)$. Define

$$h'_n(\omega) = \min_{s'_1,..,s'_m \in S} \left\{ \frac{h_n(s'_1, .., s'_m, \omega)}{f^*(s'_1, .., s'_m)} : f^*(s'_1, .., s'_m) > 0 \right\}, \ \ \omega \in S^\infty.$$

Then $h'_n \in \mathcal{F}^\ell$ by [2, Lemma 17.30]. We will show that

$$h_n(\omega) \geq f^*(\omega) \cdot \theta^m h'_n(\omega) \geq (f^* \cdot \theta^m f)(\omega) \ \text{ for each } \omega \in S^\infty. \tag{B.5}$$

Fix ω. If $f^*(\omega) = 0$, the inequality is clear. Assume $f^*(\omega) > 0$.

The first inequality in (B.5) holds because

$$\begin{aligned} f^*(\omega) \cdot \theta^m h'_n(\omega) &= f^*(s_1, ..., s_m) \cdot h'_n(s_{m+1}, ...) \\ &\leq f^*(s_1, ..., s_m) \cdot \frac{h_n(s_1, .., s_m, s_{m+1}, ...)}{f^*(s_1, .., s_m)} \\ &= h_n(s_1, .., s_m, s_{m+1}, ...) . \end{aligned}$$

For the second inequality, $f^*(s'_1, ..., s'_m) \cdot f\left(s'_{m+1}, ...\right) \leq h_n(\omega')$ for each $\omega' = (s'_1, s'_2, ...)$. Therefore,

$$f(s_{m+1}, ...) \leq \frac{h_n(s_1, s_2, ...)}{f^*(s_1, ..., s_m)}$$

whenever $f^*(s_1, ..., s_m) > 0$, and

$$f(s_{m+1}, ...) \leq \min_{s_1,..,s_m \in S} \frac{h_n(s_1, s_2, ...)}{f^*(s_1, ..., s_m)} = h'_n(s_{m+1}, ...) ,$$

which completes the proof of (B.5).

Finally, since U is monotone, $U(h_n) \geq U(f^* \cdot \theta^m h'_n) \geq U(f^* \cdot \theta^m f)$. Thus,

$$[U(h_n) \searrow U(f^* \cdot \theta^m f)] \Longrightarrow [U(f^* \cdot \theta^m h'_n) \searrow U(f^* \cdot \theta^m f)],$$

which proves Outer Regularity for U^*. ■

It is evident that U^* (or the preference that it represents) satisfies the Gilboa-Schmeidler axioms. Symmetry is satisfied because $U(f^* \cdot \theta^m f) = U(f^* \cdot (\theta^m(\pi f)))$ for any permutation π, by Symmetry for U. For Orthogonal Independence, let f, f' be nonhedging, $f, f' \in \mathcal{F}_I$ and $f^{**} \in \mathcal{F}_{I^{**}}$ with finite and disjoint I and I^{**}. Then

$$\begin{aligned} & U(f^* \cdot \theta^m [\alpha(f^{**} \cdot f) + (1-\alpha)(f^{**} \cdot f')]) \\ = \ & U(\alpha(f^* \cdot \theta^m f^{**}) \cdot \theta^m f + (1-\alpha)(f^* \cdot \theta^m f^{**}) \cdot \theta^m f') \\ = \ & \alpha U((f^* \cdot \theta^m f^{**}) \cdot \theta^m f) + (1-\alpha) U((f^* \cdot \theta^m f^{**}) \cdot \theta^m f') \\ = \ & \alpha U(f^* \cdot \theta^m (f^{**} \cdot f)) + (1-\alpha) U(f^* \cdot \theta^m (f^{**} \cdot f')) . \end{aligned}$$

This implies OI for U^*. Super-Convexity is also immediate. Conclude that $U^* \in \mathcal{U}^*$.

It remains to prove that $U^{**} \in \mathcal{U}^*$. This is more difficult because U^{**} is a difference of two functions derived from U. We show that U^{**} is suitably monotone and concave and that it satisfies Regularity. Other properties are immediate.

Lemma B.9. *The function U^{**} defined in (B.2) is monotone on $\mathcal{F}_{fin}$, that is, for all f', f in $\mathcal{F}_{fin}$,*

$$f' \geq f \Longrightarrow U^{**}(f') \geq U^{**}(f).$$

Proof. Take $f^* \in \mathcal{F}_m$ and $f, f' \in \mathcal{F}_n$. For each $\tau \in [0,1]$, there is a common minimizing measure P_τ for $f^* \cdot \theta^m(\tau f' + (1-\tau) f)$ and $\theta^m(\tau f' + (1-\tau) f)$, by OI and Lemma 4.1. Let $\varphi(P,\tau) = \int f^* \cdot \theta^m(\tau f' + (1-\tau) f)\, dP$. Then

$$U(f^* \cdot \theta^m(\tau f' + (1-\tau) f)) = \min_{P \in \mathcal{P}} \varphi(P,\tau) = \int f^* \cdot \theta^m(\tau f' + (1-\tau) f)\, dP_\tau.$$

The partial derivative with respect to τ is $\varphi_\tau(P,\tau) = \int f^* \cdot \theta^m(f' - f)\, dP$. Therefore, by [28, Theorem 2],

$$U(f^* \cdot \theta^m f') - U(f^* \cdot \theta^m f) = \int_0^1 \left[\int f^* \cdot \theta^m(f' - f)\, dP_\tau\right] d\tau.$$

Similarly,

$$U(\theta^m f') - U(\theta^m f) = \int_0^1 \left[\int \theta^m(f' - f)\, dP_\tau\right] d\tau.$$

Therefore,

$$\begin{aligned} & (1 - U(f^*))(U^{**}(f') - U^{**}(f)) \\ = \; & U(\theta^m f') - U(f^* \cdot \theta^m f') - U(\theta^m f) + U(f^* \cdot \theta^m f) \\ = \; & \int_0^1 \left[\int (1 - f^*) \cdot \theta^m(f' - f)\, dP_\tau\right] d\tau. \end{aligned}$$

Conclude that if $f' \geq f$, then $\int (1 - f^*) \cdot \theta^m(f' - f)\, dP_\tau \geq 0$ for all $\tau \in [0,1]$, and $U^{**}(f') \geq U^{**}(f)$. ■

Lemma B.10. *If $F \in \mathcal{F}_{\{1,\ldots,n\}}$ and if g^* and h^* are nonhedging, then so are*

$$g^{**} = \tfrac{1}{2}\theta^n g^* + \tfrac{1}{2} F \cdot \theta^n h^* \text{ and } h^{**} = \tfrac{1}{2}\theta^n h^* + \tfrac{1}{2} F \cdot \theta^n g^*. \tag{B.6}$$

Proof. Compute, using OI and Lemma 4.1 repeatedly, that

$$\begin{aligned} U\left(\tfrac{1}{2} g^{**} + \tfrac{1}{2} h^{**}\right) &= U\left(\left(\tfrac{1}{2}\mathbf{1} + \tfrac{1}{2}F\right) \cdot \theta^n\left(\tfrac{1}{2} g^* + \tfrac{1}{2} h^*\right)\right) \\ &= \tfrac{1}{2} U\left(\left(\tfrac{1}{2}\mathbf{1} + \tfrac{1}{2}F\right) \cdot \theta^n g^*\right) \\ &\quad + \tfrac{1}{2} U\left(\left(\tfrac{1}{2}\mathbf{1} + \tfrac{1}{2}F\right) \cdot \theta^n h^*\right) \\ &= \tfrac{1}{4} U(\theta^n g^*) + \tfrac{1}{4} U(F \cdot \theta^n g^*) \\ &\quad + \tfrac{1}{4} U(\theta^n h^*) + \tfrac{1}{4} U(F \cdot \theta^n h^*) \\ &= \tfrac{1}{2}\left[\tfrac{1}{2} U(\theta^n g^*) + \tfrac{1}{2} U(F \cdot \theta^n h^*)\right] \\ &\quad + \tfrac{1}{2}\left[\tfrac{1}{2} U(\theta^n h^*) + \tfrac{1}{2} U(F \cdot \theta^n g^*)\right] \\ &= \tfrac{1}{2} U\left(\tfrac{1}{2}\theta^n g^* + \tfrac{1}{2} F \cdot \theta^n h^*\right) \\ &\quad + \tfrac{1}{2} U\left(\tfrac{1}{2}\theta^n h^* + \tfrac{1}{2} F \cdot \theta^n g^*\right) \\ &= \tfrac{1}{2} U(g^{**}) + \tfrac{1}{2} U(h^{**}). \end{aligned}$$

■

Lemma B.11. *The function U^{**} defined in (B.2) satisfies Super-Convexity.*

Proof. Let $g^* \geq h^* \in \mathcal{F}_{\{1,...,m\}}$ be nonhedging. Since the denominator $1 - U(f^*)$ is not important, consider the function U_2 defined by the numerator. Then

$$\begin{aligned}
U_2(g^* \cdot \theta^m f) - U_2(h^* \cdot \theta^m f) &= U(g^* \cdot \theta^m f) - U(f^* \cdot \theta^n (g^* \cdot \theta^m f)) \\
&\quad - [U(h^* \cdot \theta^m f) - U(f^* \cdot \theta^n (h^* \cdot \theta^m f))] \\
&= [U(g^* \cdot \theta^m f) + U(f^* \cdot \theta^n (h^* \cdot \theta^m f))] \\
&\quad - [U(h^* \cdot \theta^m f) + U(f^* \cdot \theta^n (g^* \cdot \theta^m f))] \\
&= \left[U(g^* \cdot \theta^m f) + U\left(f^* \cdot \theta^n h^* \cdot \theta^{n+m} f\right)\right] \\
&\quad - \left[U(h^* \cdot \theta^m f) + U\left(f^* \cdot \theta^n g^* \cdot \theta^{n+m} f\right)\right] \\
\text{(by Symmetry)} &= \left[U\left(\theta^n g^* \cdot \theta^{n+m} f\right) + U\left(f^* \cdot \theta^n h^* \cdot \theta^{n+m} f\right)\right] \\
&\quad - \left[U\left(\theta^n h^* \cdot \theta^{n+m} f\right) + U\left(f^* \cdot \theta^n g^* \cdot \theta^{n+m} f\right)\right] \\
\text{(by OI)} &= 2U\left(\left(\tfrac{1}{2}\theta^n g^* + \tfrac{1}{2} f^* \cdot \theta^n h^*\right) \cdot \theta^{n+m} f\right) \\
&\quad - 2U\left(\left(\tfrac{1}{2}\theta^n h^* + \tfrac{1}{2} f^* \cdot \theta^n g^*\right) \cdot \theta^{n+m} f\right) \\
&= 2\left[U\left(g^{**} \cdot \theta^{n+m} f\right) - U\left(h^{**} \cdot \theta^{n+m} f\right)\right],
\end{aligned}$$

where g^{**} and h^{**} are defined in (B.6). Note that g^{**} and h^{**} are nonhedging by Lemma B.10. Also, $g^{**} \geq h^{**}$. Therefore, Super-Convexity for U implies that it is satisfied also by U^{**}. ■

It remains to prove regularity and also that monotonicity and concavity obtain on all of $\mathcal{F}$. For this purpose we exploit the regularity of U, as described in the following lemmas. As the surrounding arguments are routine, many details are omitted.

Let $\mathcal{F}^{\ell}_{fin} = \mathcal{F}^{\ell} \cap \mathcal{F}_{fin}$, the set of (simple) lsc acts that are finitely-based.[34]

Lemma B.12. *Let $f^* \in \mathcal{F}_{\{1,...,m\}}$. Then, for any $f', f \in \mathcal{F}$ and $\alpha \in [0,1]$,*

$$U(f^* \cdot \theta^m (\alpha f + (1-\alpha) f')) = \inf_{\substack{f \leq h \in \mathcal{F}^{\ell} \\ f' \leq h' \in \mathcal{F}^{\ell}}} U(f^* \cdot \theta^m (\alpha h + (1-\alpha) h')).$$

Proof. $\mathcal{P}$ denotes the set of priors corresponding to U. Note that

$$\begin{aligned}
&U(\alpha (f^* \cdot \theta^m f) + (1-\alpha)(f^* \cdot \theta^m f')) \\
&= \inf_{P \in \mathcal{P}} \left[\alpha \int (f^* \cdot \theta^m f)\, dP + (1-\alpha) \int (f^* \cdot \theta^m f')\, dP\right]
\end{aligned}$$

[34]Since S is finite, every finitely-based act is continuous, hence lsc. However, we use the notation $\mathcal{F}^{\ell}_{fin}$ in order to emphasize that we are using the lower semi-continuity of such acts, which would be important in any future generalization to infinite S.

$$
\begin{aligned}
&= \inf_{P\in\mathcal{P}} \left[\alpha \inf_{f\le h\in\mathcal{F}^\ell} \int (f^*\cdot\theta^m h)\, dP + (1-\alpha) \inf_{f'\le h'\in\mathcal{F}^\ell} \int (f^*\cdot\theta^m h')\, dP\right] \\
&= \inf_{P\in\mathcal{P}} \inf_{\substack{f\le h\in\mathcal{F}^\ell \\ f'\le h'\in\mathcal{F}^\ell}} \int [\alpha (f^*\cdot\theta^m h) + (1-\alpha)(f^*\cdot\theta^m h')]\, dP \\
&= \inf_{\substack{f\le h\in\mathcal{F}^\ell \\ f'\le h'\in\mathcal{F}^\ell}} \inf_{P\in\mathcal{P}} \int [\alpha (f^*\cdot\theta^m h) + (1-\alpha)(f^*\cdot\theta^m h')]\, dP \\
&= \inf_{\substack{f\le h\in\mathcal{F}^\ell \\ f'\le h'\in\mathcal{F}^\ell}} U(\alpha (f^*\cdot\theta^m h) + (1-\alpha)(f^*\cdot\theta^m h')) .
\end{aligned}
$$

The second equality follows because $f \longmapsto \int f dP$, for $P \in \Delta(S^\infty)$, is monotone and satisfies Regularity; hence Lemma B.8 implies $\int f^* \cdot \theta^m f dP =$ $\inf_{f\le h\in\mathcal{F}^\ell} \int f^*\cdot\theta^m h dP$. ■

Lemma B.13. *Let* $f^* \in \mathcal{F}_{\{1,\dots,m\}}$.

(a) For any $h\in\mathcal{F}^\ell$, *there exist* $h_n \in \mathcal{F}^\ell_{fin}$ *such that* $h_n \le h$,

$$U(h_n) \nearrow U(h) \text{ and } U(f^*\cdot\theta^m h_n) \nearrow U(f^*\cdot\theta^m h).$$

(b) For any $f', f\in\mathcal{F}$ *and* $\alpha\in[0,1]$, *there exist* $h_n, h'_n \in \mathcal{F}^\ell$ *such that*

$$
\begin{aligned}
& f\le h_n,\ f\le h'_n\ ,\ U(h_n)\searrow U(f),\ U(h'_n)\searrow U(f'), \\
& U(f^*\cdot\theta^m h_n)\searrow U(f^*\cdot\theta^m h)\ ,\ U(f^*\cdot\theta^m h'_n)\searrow U(f^*\cdot\theta^m h')\ , \\
& U(\alpha h_n + (1-\alpha) h'_n) \searrow U(\alpha f + (1-\alpha) f') \text{ and} \\
& U(f^*\cdot\theta^m(\alpha h_n + (1-\alpha) h'_n)) \searrow U(f^*\cdot\theta^m(\alpha f + (1-\alpha) f')) .
\end{aligned}
$$

Proof. (a) By Inner Regularity, there is a sequence $g_n \in \mathcal{F}^u$ such that $g_n \le h$ and $U(g_n) \nearrow U(h)$. By [17, Lemma D.3], there exists $h'_n \in \mathcal{F}^\ell_{fin}$ such that $g_n \le h'_n \le h$. Then $U(h'_n) \nearrow U(h)$. Similarly, by the regularity established in Lemma B.8, there exist $h''_n \in \mathcal{F}^\ell_{fin}$ such that $h''_n \le h$ and $U(f^*\cdot\theta^m h''_n) \nearrow U(f^*\cdot\theta^m h)$. Define $h_n = \max\{h'_n, h''_n\}$ and h_n does the job.

(b) By Regularity of U, there is $\widehat{h_n} \in \mathcal{F}^\ell$ such that $\widehat{h_n} \ge f$ and $U\left(\widehat{h_n}\right) \searrow U(f)$. By the regularity established in Lemma B.8, there exist $\widehat{\widehat{h_n}} \in \mathcal{F}^\ell$ such that $\widehat{\widehat{h_n}} \ge f$ and $U(f^*\cdot\theta^m\widehat{\widehat{h_n}}) \searrow U(f^*\cdot\theta^m f)$. Define $h_n\in\mathcal{F}^\ell$ by

$$h_n = \min\left\{\widehat{h_n}, \widehat{\widehat{h_n}}\right\}.$$

Then $h_n\in\mathcal{F}^\ell$, $h_n \ge f$ and

$$U(h_n)\searrow U(f) \text{ and } U(f^*\cdot\theta^m h_n)\searrow U(f^*\cdot\theta^m f).$$

The preceding argument is readily extended to prove the remainder of (b), when combined with Lemma B.12. ■

We can finally complete the proof of Lemma B.6.

Monotonicity: $U^{**}(f') \geq U^{**}(f)$ if $f' \geq f$ and $f', f \in \mathcal{F}$. By Lemma B.9, this is true if f' and f are finitely-based. The inequality is readily extended to all simple lsc acts, and then to arbitrary acts, by using Lemma B.13.

Regularity: Since U^{**} is increasing, $U^{**}(f) \leq \inf\left\{U^{**}(h) : h \geq f, h \in \mathcal{F}^{\ell}\right\}$. Lemma B.13 implies equality, which proves Outer Regularity. Inner Regularity can be shown similarly.

Concavity: We have to show that

$$U^{**}(\alpha f + (1-\alpha) f') \geq \alpha U^{**}(f) + (1-\alpha) U^{**}(f'), \text{ for all } f', f \in \mathcal{F}.$$

For finitely-based f' and f, the inequality follows from Lemma B.11. It is readily extended to all simple lsc acts, and then to arbitrary acts, by using Lemma B.13.

We offer a remark related to the proof. Above we showed that every extreme point of $\mathcal{U}^*$ lies in $\mathcal{V}$. In fact, we can prove, using the representation, that the other direction is also true.

Lemma B.14. *$\mathcal{V}$ is the set of all extreme points of $\mathcal{U}^*$.*

Proof. Let $U \in \mathcal{V}$ and show that U is an extreme point of $\mathcal{U}^*$.

The proof of Theorem 5.2, specifically, application of Choquet's Theorem, implies that $U(f) = \int V(f)\, d\mu(V)$ for some μ that is supported by the set of extreme points of $\mathcal{U}^*$ (and not only by its superset $\mathcal{V}$). Therefore, for $f \in \mathcal{F}_{\{1,\ldots,m\}}$,

$$\begin{aligned}\left[\int V(f)\, d\mu(V)\right]^2 &= [U(f)]^2 = U(f \cdot \theta^m f) \\ &= \int V(f \cdot \theta^m f)\, d\mu(V) = \int [V(f)]^2\, d\mu(V).\end{aligned}$$

But $\left[\int V(f)\, d\mu(V)\right]^2 = \int [V(f)]^2\, d\mu(V)$ if and only if

$$V(f) \text{ is constant } \mu\text{-}a.s.[V].$$

The exceptional set depends on f. But since $\mathcal{F}_1$ is separable, there exists a μ-null set of V's that works for all acts. Conclude that $a.s.$-$\mu[V]$, $V(\cdot) = U(\cdot)$ on $\mathcal{F}_{\{1,\ldots,m\}}$. Since this is true for any m, the equality holds $a.s.$ on all of $\mathcal{F}$ by the generalized Kolmogorov extension theorem [17, Theorem D.2]. Thus, μ is degenerate and U is an extreme point of $\mathcal{U}^*$. ■

B.4. Uniqueness

Let μ' and μ, Borel measures on the compact metric space $\mathcal{V}$, satisfy

$$\int V(f)\, d\mu' = \int V(f)\, d\mu \text{ for all } f \in \mathcal{F}.$$

We show that

$$\mu' = \mu.$$

Each finitely-based act f induces (by Lemma B.4) the continuous map $\widehat{f}:\mathcal{V}\rightarrow[0,1]$, given by

$$\widehat{f}\left(V\right)=V\left(f\right).$$

Let $\widehat{\mathcal{F}}_{fin}$ be the set of all such maps and $\mathcal{A}=sp\left(\widehat{\mathcal{F}}_{fin}\right)$, the linear span of $\widehat{\mathcal{F}}_{fin}$ within $C\left(\mathcal{V}\right)$, the set of continuous real-valued functions on $\mathcal{V}$. Then,

$$\int\widehat{f}(V)d\mu'=\int\widehat{f}(V)d\mu\text{ for all }\widehat{f}\in\widehat{\mathcal{F}}_{fin}.$$

This equality extends also to the linear span:

$$\int\phi(V)d\mu'=\int\phi(V)d\mu\text{ for all }\phi\in\mathcal{A}.$$

It is enough to show that

$$\int\phi(V)d\mu'=\int\phi(V)d\mu\text{ for all }\phi\in C\left(\mathcal{V}\right).\tag{B.7}$$

We do this by verifying the conditions of the Stone-Weierstrass Theorem, which implies that $\mathcal{A}$ is sup-norm dense in $C\left(\mathcal{V}\right)$, and hence also (B.7).

Obviously $\mathcal{A}$ contains the constant functions and it separates points; in fact, since every IID utility is regular, if $V'\neq V$, then $\phi\left(V'\right)\neq\phi\left(V\right)$ for some $\phi\in\widehat{\mathcal{F}}_{fin}\subset\mathcal{A}$. We need only show that

$$\phi',\phi\in\mathcal{A}\Longrightarrow\phi'\phi\in\mathcal{A},$$

which follows from Steps 1 and 2.

Step 1. Any finite linear combination of elements in $\widehat{\mathcal{F}}_{fin}$ can be expressed as a linear combination of two such elements, that is,

$$\Sigma_i a_i\widehat{f}_i=\kappa\widehat{h}-\kappa'\widehat{h}'.\tag{B.8}$$

Clearly,

$$\left(\Sigma_i a_i\widehat{f}_i\right)\left(V\right)=\Sigma_i a_i\widehat{f}_i\left(V\right)=\Sigma_i a_i V(f_i).$$

Suppose that every a_i is positive. We can shift each of the acts f_i so that they are mutually orthogonal and V is additive over them (since every IID utility satisfies OI). Because weights may not sum to 1, we obtain $\kappa V\left(h\right)$ for some finitely based act h and $\kappa>0$, that is,

$$\Sigma_i a_i\widehat{f}_i=\kappa\widehat{h}.$$

If one or more of the coefficients a_i is negative, then one can collect those acts having similarly signed weights, and derive (B.8).

Step 2. Verify that $(a\widehat{f}+b\widehat{g})(a'\widehat{f}'+b'\widehat{g}')\in\mathcal{A}$:

$$\begin{aligned}
&[(a\widehat{f}+b\widehat{g})(V)]\;[(a'\widehat{f}'+b'\widehat{g}')\left(V\right)]\\
=\;&[aV\left(f\right)+bV\left(g\right)]\;\;[a'V\left(f'\right)+b'V\left(g'\right)]\\
=\;&aa'V\left(f\cdot\theta^n f'\right)+ab'V\left(f\cdot\theta^n g'\right)+ba'V\left(g\cdot\theta^n f'\right)+bb'V\left(g\cdot\theta^n g'\right)\\
=\;&\left(aa'(\widehat{f\cdot\theta^n f'})+ab'(\widehat{f\cdot\theta^n g'})+ba'(\widehat{g\cdot\theta^n f'})+bb'(\widehat{g\cdot\theta^n g'})\right)\left(V\right),
\end{aligned}$$

where n is large enough so that all paired acts are orthogonal to one another. The last equality is derived by shifting each of the product acts so that they are mutually orthogonal, so that V is additive over them, and then applying shift invariance. Thus (B.8) implies

$$(a\widehat{f}+b\widehat{g})(a'\widehat{f}'+b'\widehat{g}')=\kappa\widehat{h}-\kappa'\widehat{h}'\in\mathcal{A}.$$

■

C. Appendix: Proof of Theorem 7.2

Step 1: $V(g\cdot f)=V(g)V(f)$ for all $g\in\mathcal{F}_{\{1,\ldots,m\}}$ and $f\in\mathcal{F}_{\{m+1,m+2,\ldots\}}$.

The equality is true by (5.1) if f is finitely-based. Extend it to all acts f indicated by applying Regularity.

Step 2: Fix $A\in\Sigma^{tail}$ and define, (where A denotes $\mathbf{1}_A$ and so on),

$$\mathcal{B}=\left\{B\in\Sigma:V\left(\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right)=V\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right\}.$$

Then $\mathcal{B}$ is a monotone class.

(a) Assume $B_n\in\mathcal{B}$, $B_n\nearrow B$ and show $B\in\mathcal{B}$, that is,

$$V\left(\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right)=V\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}A+\tfrac{1}{2}\right).$$

Let $C_n=B\backslash B_n\searrow\varnothing$, and define, for a fixed tail event A',

$$\begin{aligned}f&=\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right),\\ f_n&=\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\text{ and}\\ g_n&=\left(\tfrac{1}{4},C_n;f_n,\Omega\backslash C_n\right).\end{aligned}$$

(a.i) If $s\in C_n$, then $s\notin B_n$ and $f_n(s)=\frac{1}{2}\left(\frac{1}{2}A'+\frac{1}{2}\right)(s)\in\left\{\frac{1}{4},\frac{1}{2}\right\}$.
(a.ii) By (a.i), $g_n\leq f_n$. Therefore, $V(g_n)\leq V(f_n)$.
(a.iii) $f_n(s)\neq f(s)\Longrightarrow$

$$\left[\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right](s)\neq\left[\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right](s)$$

$$\Longrightarrow\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)(s)\neq\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)(s)\Longrightarrow s\in B\backslash B_n=C_n.$$

Therefore, $s\notin C_n\Longrightarrow f_n(s)=f(s)$.
(a.iv) $g_n=\left(\frac{1}{4},C_n;f,\Omega\backslash C_n\right)$. This is clear by (a.iii).
(a.v) By Monotone Continuity, for any $\epsilon>0$, there exists N such that $V(g_N)>V(f)-\epsilon$. Therefore, by (a.ii), $V(f_N)>V(f)-\epsilon$. But $f_n\overset{n}{\nearrow}f$. Conclude that

$$V(f_n)\nearrow V(f).\tag{C.1}$$

We can now complete the proof of (a) and show that $B\in\mathcal{B}$: (C.1) implies that

$$\lim V\left(\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right)=V\left(\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right)$$

for all $A' \in \Sigma^{tail}$. Thus

$$\begin{aligned} V\left(\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right) &= \lim V\left(\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right) \quad (\text{set } A'=A) \\ &= \lim V\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}A+\tfrac{1}{2}\right) \quad (\text{since } B_n\in\mathcal{B}) \\ &= V\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}A+\tfrac{1}{2}\right). \quad (\text{set } A'=\Omega) \end{aligned}$$

(b) Assume $B_n \in \mathcal{B}$, $B_n \searrow B$ and show that $B \in \mathcal{B}$. The argument is similar to that in (a). We provide an outline for completeness.

Let $C_n = B_n \backslash B \searrow \varnothing$, and define, for a fixed tail event A',

$$\begin{aligned} f &= \left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right), \\ f_n &= \left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right) \text{ and} \\ g_n &= \left(1, C_n; f_n, \Omega\backslash C_n\right). \end{aligned}$$

(b.i) If $s \in C_n$, then $s \in B_n$ and $f_n(s) = \left(\frac{1}{2}A'+\frac{1}{2}\right)(s) \in \left\{\frac{1}{2}, 1\right\}$.
(b.ii) By (b.i), $g_n \geq f_n$. Therefore, $V(g_n) \geq V(f_n)$.
(b.iii) $f_n(s) \neq f(s) \Longrightarrow$

$$\left[\left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right](s) \neq \left[\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A'+\tfrac{1}{2}\right)\right](s)$$

$$\Longrightarrow \left(\tfrac{1}{2}B_n+\tfrac{1}{2}\right)(s) \neq \left(\tfrac{1}{2}B+\tfrac{1}{2}\right)(s) \Longrightarrow s \in B_n\backslash B = C_n.$$

Therefore, $s \notin C_n \Longrightarrow f_n(s) = f(s)$.
(b.iv) $g_n = (1, C_n; f, \Omega\backslash C_n)$. This is clear by (b.iii).
(b.v) $V(f_n) \searrow V(f)$.

The rest of the argument is exactly as in (a).

Step 3: By Step 1, $\cup_m \Sigma_{\{1,\ldots,m\}} \subset \mathcal{B}$. Thus the Monotone Class Lemma [2, p. 137] implies that $\mathcal{B} = \Sigma$, that is, for all $A \in \Sigma^{tail}$ and $B \in \Sigma$,

$$V\left(\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}A+\tfrac{1}{2}\right)\right) = V\left(\tfrac{1}{2}B+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}A+\tfrac{1}{2}\right).$$

In the same way we can show that, for all $A \in \Sigma^{tail}$ and $B \in \Sigma$,

$$V(1_A \cdot 1_B) = V(1_{A\cap B}) = V(A\cap B) = V(A)V(B). \tag{C.2}$$

The rest of the proof uses these properties and not Monotone Continuity directly.

Step 4: Apply Step 3 to two tail events A and B to derive

$$\begin{aligned} V\left(\tfrac{1}{4}1_B\cdot 1_A+\tfrac{1}{4}1_A+\tfrac{1}{4}1_B+\tfrac{1}{4}\right) &= V\left(\left(\tfrac{1}{2}1_B+\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}1_A+\tfrac{1}{2}\right)\right) \\ &= V\left(\tfrac{1}{2}1_B+\tfrac{1}{2}\right)V\left(\tfrac{1}{2}1_A+\tfrac{1}{2}\right) \\ &= \tfrac{1}{4}\left[V(A)V(B)+V(A)+V(B)+1\right] \\ &= \tfrac{1}{4}\left[V(A\cap B)+V(A)+V(B)+1\right]. \end{aligned}$$

By Lemma 2.2,

$$V\left(\tfrac{1}{2}1_A+\tfrac{1}{2}1_B\right) = \tfrac{1}{2}V(A)+\tfrac{1}{2}V(B).$$

Step 5: $A \longmapsto V(A)$ defines a finitely additive 0-1 valued measure (or charge) on Σ^{tail}: The 0-1 property follows from (C.2). For disjoint $A, B \in \Sigma^{tail}$, by Step 4,

$$\begin{aligned} V(A \cup B) &= V(1_{A \cup B}) = V(1_A + 1_B) \\ &= 2V\left(\tfrac{1}{2}1_A + \tfrac{1}{2}1_B\right) = V(A) + V(B). \end{aligned}$$

Step 6: Let $\mathcal{P}$ be the set of priors corresponding to V. For $A \in \Sigma^{tail}$, $V(A) + V(\Omega \backslash A) = 1$. Thus, $V(A) = 0 \Longrightarrow V(\Omega \backslash A) = 1$. Further, $V(A) = 1 \Longrightarrow P(A) = 1$ for all $P \in \mathcal{P}$. Since $V(A) = 0$ or 1, it follows that

$$\{P(A) : P \in \mathcal{P}\} = \{0\} \text{ or } \{1\}.$$

Step 7: For each $f \in \mathcal{F}_1$, there is an exchangeable measure P^* that is minimizing for f. To see this, note that

$$\begin{aligned} V\left(\tfrac{1}{4}f \cdot \theta f + \tfrac{1}{4}f + \tfrac{1}{4}\theta f + \tfrac{1}{4}\right) &= V\left(\left(\tfrac{1}{2}f + \tfrac{1}{2}\right) \cdot \left(\tfrac{1}{2}\theta f + \tfrac{1}{2}\right)\right) \\ &= V\left(\tfrac{1}{2}f + \tfrac{1}{2}\right) V\left(\tfrac{1}{2}\theta f + \tfrac{1}{2}\right) \\ &= \tfrac{1}{4}\left[V(f \cdot \theta f) + V(f) + V(\theta f) + 1\right]. \end{aligned}$$

By Lemma 2.2, there is a common minimizing measure P for f and θf. Let π be the permutation that switches experiments 1 and 2. Then, using Symmetry,

$$(\pi P) f = P(\pi f) = P(\theta f) = V(\theta f) = V(f).$$

Therefore, P and πP are both minimizing for f. Finally, $P^1 \equiv \frac{1}{2}P + \frac{1}{2}\pi P$ is also minimizing (it lies in $\mathcal{P}$ because $\mathcal{P}$ is convex) and it satisfies $\pi P^1 = P^1$.

Apply a similar argument to $\left(\frac{1}{2}f + \frac{1}{2}\right) \cdot \left(\frac{1}{2}\theta f + \frac{1}{2}\right) \ldots \cdot \left(\frac{1}{2}\theta^n f + \frac{1}{2}\right)$ to deduce that there is a common minimizing measure P^n for $\{f, \theta f, ..., \theta^n f\}$ that satisfies $\pi P^n = P^n$ for all $\pi \in \Pi^n$, the set of permutations on $\{1, ..., n\}$. Since $\mathcal{P}$ is compact, wlog (after relabelling), $P^n \rightarrow P^* \in \mathcal{P}$. Then P^* is exchangeable and minimizing for f.

Step 8: The measure P^* in Step 8 is i.i.d.: By Step 6, P^* is 0-1 valued on Σ^{tail}. But, using the de Finetti Theorem, it is straightforward to show that the only exchangeable measures with this property are i.i.d. measures.

Step 9. $V(\alpha f' + (1-\alpha) f) = \alpha V(f') + (1-\alpha) V(f)$, for all $f', f \in \mathcal{F}_1$.

Take i.i.d. measures P' for f' and P for f. Since both P' and P are i.i.d. measures, and they agree on tail events (Step 6), they must coincide. Thus, there is a common minimizing measure for f' and f.

Step 10. $V(\alpha f' + (1-\alpha) f) = \alpha V(f') + (1-\alpha) V(f)$, for all $f', f \in \mathcal{F}$.

For any n, view S^n as corresponding to one experiment and repeat the above to derive additivity for all $f', f \in \mathcal{F}_{\{1,\ldots,n\}}$. Finally, apply Regularity to extend additivity to all acts. ■

D. Appendix: Proofs for Updating

Proof of Theorem 6.1: Prove sufficiency of the axioms. We prove (6.1) for $n = 1$; the general argument is similar.

We use Proposition 1 in de Meyer and Mongin [27], for which the main step is to show that D is convex, where

$$D = \left\{ \left(U(f), U_1(f \mid s_1) \right)_{s_1 \in S_1} : f \in \mathcal{F}_{>1} \right\} \subset \mathbb{R}^{S+1}.$$

A preliminary result concerns shifted acts. Recall that θ is the *shift* operator, so that, for any act,

$$(\theta f)(s_1, s_2, s_3, ...) = f(s_2, s_3, ..);$$

θ^n denotes the n-fold replication of θ. Symmetry implies also indifference to shifts, that is, $\theta f \sim f$ for all acts f (see Section 4.2).

Now let $x, y \in D$,

$$x = \left(U(f), U_1(f \mid s_1) \right)_{s_1 \in S_1} \text{ and } y = \left(U(g), U_1(g \mid s_1) \right)_{s_1 \in S_1},$$

and prove that $\alpha x + (1-\alpha) y \in D$. Suppose first that f and g finitely-based. Then there exists N large enough so that f and the shifted act $\theta^N g$ are orthogonal, that is, they depend on disjoint sets of experiments. For such an N, because each utility function satisfies OI and shift-invariance,

$$\begin{aligned} \alpha x + (1-\alpha) y &= \alpha \left(U(f), U_1(f \mid s_1) \right)_{s_1 \in S_1} + (1-\alpha) \left(U(g), U_1(g \mid s_1) \right)_{s_1 \in S_1} \\ &= \alpha \left(U(f), U_1(f \mid s_1) \right)_{s_1 \in S_1} + (1-\alpha) \left(U\left(\theta^N g\right), U_1\left(\theta^N g \mid s_1\right) \right)_{s_1 \in S_1} \\ &= \left(U\left(\alpha f + (1-\alpha)\theta^N g\right), U_1\left(\alpha f + (1-\alpha)\theta^N g \mid s_1\right) \right)_{s_1 \in S_1} \in D, \end{aligned}$$

where the last equality follows from OI. Finally, the preceding can be extended to general (not only finitely-based) acts f and g by Regularity.

The other conditions in Proposition 1 of de Meyer and Mongin [27] are readily verified.[35] Therefore, there exist positive numbers $a_{s_1} > 0$ such that

$$U(f) = \Sigma_{s_1} a_{s_1} U_1(f \mid s_1), \; f \in \mathcal{F}_{>1}.$$

Since $U(p) = U_1(p \mid s_1) = p$ for all (constant acts) p, it follows that $\Sigma_{s_1} a_{s_1} = 1$.

Deduce that, for all $f \in \mathcal{F}_{>1}$,

$$\int V(f)\, d\mu(V) = \Sigma_{s_1} a_{s_1} \int V(f)\, d\mu_{s_1}(V) = \int V(f) \left(\Sigma_{s_1} a_{s_1} d\mu_{s_1}(V) \right).$$

By uniqueness of the representing measure,

$$\mu(\cdot) = \Sigma_{s_1} a_{s_1} \mu_{s_1}(\cdot).$$

[35] De Meyer and Mongin's condition (C) is satisfied here because $U(p) = U_1(p \mid s_1) = p$ for all s_1 and $0 \le p \le 1$. Therefore, WDC implies their condition P_4, and the Proposition's conclusion follows.

Because $a_{s_1} > 0$ for each s_1, it follows that $\mu_{s_1} << \mu$, and

$$1 = \Sigma_{s_1} a_{s_1} \left(d\mu_{s_1} (\cdot) / d\mu (\cdot) \right).$$

Equation (6.1) is satisfied for $n = 1$ if

$$L_1 (s_1 \mid V) = a_{s_1} \left(d\mu_{s_1} (V) / d\mu (V) \right).$$

Similarly for $n > 1$.

Argue similarly for every n to obtain a family $\{L_n (\cdot \mid V)\}$ of conditional one-step-ahead likelihoods. These can be combined in the standard way to yield a unique likelihood function $L (\cdot \mid V)$ on Ω. ■

Proof of Proposition 6.3: (i) We adapt a result of Doob as described in LeCam and Yang [26, Propositions 2,3, p. 243]. For simplicity, consider the special case of coin-tossing.

Because each $L (\cdot \mid \mathcal{I})$ is exchangeable, $\lim \Psi_n (\omega)$ exists $L (\cdot \mid \mathcal{I})$-a.s., and, for any interval $I \subset [0, 1]$,

$$\lambda_{\mathcal{I}} (I) = L (\cdot \mid \mathcal{I}) \left(\{\omega : \lim \Psi_n (\omega) \in I\} \right). \tag{D.1}$$

Since $\lambda_{\mathcal{I}}$ has support in $\mathcal{I}$, $\lambda_{\mathcal{I}} (\mathcal{I}) = 1$. Because intervals are disjoint, for each ω, there is at most one $\mathcal{I}$ such that $\lim \Psi_n (\omega) \in \mathcal{I}$. Define $F : \Omega \rightarrow Supp (\mu)$, by

$$F (\omega) = \mathcal{I}, \text{ if } \lim \Psi_n (\omega) \in \mathcal{I},$$

and define $F (\omega) = \overline{\mathcal{I}}$, with $\overline{\mathcal{I}}$ an arbitrary fixed interval in the support of μ, if $\lim \Psi_n (\omega) \notin \cup_{Supp(\mu)} \mathcal{I}$. Then,

$$\int_{Supp(\mu)} \int_{\Omega} \mid \mathcal{I} - F (\omega) \mid dL (\omega \mid \mathcal{I}) \, d\mu (\mathcal{I}) = 0,$$

which establishes the condition in Le Cam and Yang [26, Proposition 2]. Their Proposition 3 completes the proof.

(ii) Define $F : \Omega \rightarrow \{\mathcal{I}, p\}$, by $F (\omega) = p$ if $\lim \Psi_n (\omega) = p$, and $= \mathcal{I}$ otherwise. Then

$$\int_{\Omega} \mid \mathcal{I} - F (\omega) \mid dL (\omega \mid \mathcal{I}) = 0, \text{ and}$$

$$\int_{\Omega} \mid p - F (\omega) \mid dp^{\infty} (\omega) = 0.$$

The former is valid because $L (\{\omega : \lim \Psi_n (\omega) = p\} \mid \mathcal{I}) = \lambda_{\mathcal{I}} (\{p\}) = 0$. Thus Le Cam and Yang [26, Proposition 3] completes the proof. ■

References

[1] D. Acemoglu, V. Chernozhukov and M. Yildiz, Fragility of asymptotic agreement under Bayesian learning, 2008.

[2] C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis*, 3rd edition, Springer, 2006.

[3] K. Arrow, *Essays in the Theory of Risk-Bearing*, North-Holland, 1970.

[4] J.M. Bernardo, The concept of exchangeability and its applications, *Far East J. Math. Sci.* 4 (1996), 111-121.

[5] W. Brock and S.N. Durlauf, Growth empirics and reality, *The World Bank Review* 15 (2001), 229-272.

[6] A. Castaldo, F. Maccheroni and M. Marinacci, Random correspondences as bundles of random variables, *Sankhya* (Series A) 80 (2004), 409-427.

[7] A. Chateauneuf, F. Maccheroni, M. Marinacci and J.M. Tallon, Monotone continuous multiple priors, *Econ. Theory* 26 (2005), 973-982.

[8] Y.C. Chen, Universality of the Epstein-Wang type structure, *Games and Econ. Behav.*, forthcoming.

[9] G. de Cooman and E. Miranda, Symmetry of models versus models of symmetry, in *Probability and Inference: Essays in Honor of Henry E. Kyburg Jr.*, W.L. Harper and G.R. Wheeler eds., King's College Publications, London, 2007, pp. 67-149.

[10] G. de Cooman, E. Quaeghebeur and E. Miranda, Exchangeable lower previsions, *Bernoulli* 15 (2009), 721-735.

[11] B. De Finetti, La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincare* 7 (1937), 1-68. English translation in *Studies in Subjective Probability*, 2nd edition, H.E. Kyburg and H.E. Smokler eds., Krieger Publishing, Huntington NY, 1980, pp. 53-118.

[12] L.E. Dubins, On Lebesgue-like extensions of finitely additive measures, *Ann. Prob.* 2 (1974), 456-463.

[13] L.E. Dubins, Towards characterizing the set of ergodic probabilities, pp. 61-74 in *Exchangeability in Probability and Statistics,* G. Koch and F. Spizzichino eds., North Holland, Amsterdam, 1982.

[14] N. Dunford and J.T. Schwartz (1958): *Linear Operators Part I: General Theory*, Wiley.

[15] L.G. Epstein and M. Schneider, IID: independently and indistinguishably distributed, *J. Econ. Theory* 113 (2003), 32-50.

[16] L.G. Epstein and M. Schneider, Learning under ambiguity, *Rev. Ec. Studies* 74 (2007), 1275-1303.

[17] L.G. Epstein and T. Wang, 'Beliefs about beliefs' without probabilities, *Econometrica* 64 (1995), 1343-1373.

[18] P. Ghirardato, On independence for non-additive measure, with a Fubini theorem, *J. Econ. Theory* 73 (1997), 261-291.

[19] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique priors, *J. Math. Econ.* 18 (1989), 141-153.

[20] P. Hall and C.C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, 1980.

[21] J.C. Harsanyi, Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *J. Pol. Econ.* 63 (1955), 309-321.

[22] E. Hendon, H.J. Jacobson. B. Sloth and T. Tranaes, The product of capacities and belief functions, *Math. Soc. Sc.* 32 (1996), 95-108.

[23] E. Hewitt and L.J. Savage, Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.* 80 (1955), 470-501.

[24] O. Kallenberg, *Foundations of Modern Probability*, Springer, 1997.

[25] D.M. Kreps, *Notes on the Theory of Choice*, Westview, 1988.

[26] L. Le Cam and G.L. Yang, *Asymptotics in Statistics*, 2nd ed., Springer, 2000.

[27] B. de Meyer and P. Mongin, A note on affine aggregation, *Ec. Letters* 47 (1995), 177-183.

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

[29] P. Mongin, Consistent Bayesian aggregation, *J. Econ. Theory* 66 (1995), 313-351.

[30] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Springer Verlag, Berlin, 1989.

[31] F. Philippe, G. Debs and J.Y. Jaffray, Decision making with monotone lower probabilities of infinite order, *Math. Oper. Res.* 24 (1999), 767-784.

[32] E. Regazzini, Observability and probabilistic limit theorems, in *Proceedings of the International Conference "The Notion of Event in Probabilistic Epistemology,"* Trieste, 27-29 Maggio 1996.

[33] L. Savage, *The Foundations of Statistics*, Dover, 1972.

[34] M.J. Schervish, *Theory of Statistics*, Springer, 1995.

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

[36] F. Terkelsen, Some minmax theorems, *Math. Scand.* 31 (1972), 405-413.

[37] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, 1991.

[38] P. Walley and T.L. Fine, Towards a frequentist theory of upper and lower probability, *Ann. Statist.* 10 (1982), 741-761.