# SYMMETRY, AMBIGUITY AND FREQUENCIES*

Larry G. Epstein Kyoungwon Seo

May 5, 2009

**Abstract**

We model 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, and the other that is poorly understood and thought to be unrelated across experiments (the "error term"). Consequently, there is incomplete confidence that experiments are identical. We argue that a Bayesian model cannot capture the above, but that belief function utility can. Our formal contribution is to generalize the de Finetti Theorem on exchangeability to a framework where beliefs are represented by belief functions. Moreover, this is done while extending the scope of the bridge provided by de Finetti between subjectivist and frequentist approaches. In particular, a model of updating is provided.

*Boston University, lepstein@bu.edu and Department of Managerial Economics and Decision Sciences, Northwestern University, k-seo@kellogg.northwestern.edu. We have benefitted from discussions with Drew Fudenberg, Asen Kochov, Fabio Maccheroni and Jianjun Miao. First version: April 18, 2009.

# 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 various tossers 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, given the poor understanding of the error term, a sophisticated individual might very well admit the possibility that the experiments may differ in some way. Thus we refer to experiments as being *indistinguishable* but *not* necessarily *identical*.[1]

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 [6, p. 231] argue that

[1] This terminological distinction was introduced in Epstein and Schneider [13]. Another way to describe the distinction, due to Walley [41], is between symmetry of evidence, which we assume, and evidence of symmetry, which we assume is lacking, or at least, is not overwhelming. See the concluding section for more on related literature.

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 [16, 23]. There is a countable infinity of experiments, indexed by the set $\mathbb{N} = \{1, 2, ...\}$. Each experiment yields an outcome in the set $S$ (technical details are suppressed until later). Thus $\Omega = S^{\infty}$ is the set of all possible sample paths. A probability measure $P$ on $\Omega$ is *exchangeable* if

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

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

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

where, for any probability measure $\ell$ on $S$ (written $\ell \in \Delta(S)$), $\ell^{\infty}$ denotes the corresponding i.i.d. product measure on $\Omega$.[2]

[2] 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.

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 $\ell$ describes any single experiment. However, conditional on any $\ell$ in the support of $\mu$, it is the i.i.d. product $\ell^{\infty}$ that describes the implied probability law on $\Omega$. This suggests that there is no room in the model to accommodate a concern with experiments not being identical. In Section 3, 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.[3] Thus we propose a model that generalizes the exchangeable Bayesian model by suitably relaxing the Independence axiom.

Specifically, we consider preference on a domain of (Anscombe-Aumann) acts that conforms to Choquet expected utility where the capacity is a belief function - we call this model *belief function utility*.[4] Using the latter as the basic framework, we then impose two further axioms - Symmetry (corresponding to exchangeability) and Orthogonal Independence (relaxing the Independence axiom). These axioms are shown (Theorem 4.1) to characterize the following representation that extends (1.1):

$$\nu\left(\cdot\right)=\int_{Bel(S)}\theta^{\infty}\left(\cdot\right)d\mu\left(\theta\right), \tag{1.2}$$

where $\nu$ is a belief function on $\Omega$, $Bel\left(S\right)$ denotes the set of all belief functions on $S$, $\mu$ is a probability measure on $Bel\left(S\right)$, and $\theta^{\infty}$ denotes a suitable "i.i.d. product" of the belief function $\theta$.

At an informal level, the representation captures "indistinguishable but not identical" in the following way. Consider our 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. (For any $\theta\in Bel\left(S\right)$

[3] Walley [41, Ch. 9] argues that the Bayesian model cannot accommodate both symmetry of evidence and an absence of overwhelming evidence of symmetry. Brock and Durlauf's [6] 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.

[4] Belief functions and the corresponding utility functions are described in Section 2.

appearing in (1.2), the interval is $[\theta(H), 1-\theta(T)]$). 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.

There is intuition for using ambiguity averse preferences, motivated by the Ellsberg Paradox, as a framework. If the individual is not confident that the experiments are identical, then presumably there are features of each single experiment that she does not understand well. Belief function utility is particularly appealing because it is a special case of both Choquet expected utility (Schmeidler [37]) and multiple-priors utility (Gilboa and Schmeidler [20]), and thus is "close" to the benchmark expected utility model, and because it admits an epistemic rationale due to Dempster [9] and Shafer [38] (see (2.3) and the surrounding discussion).

### 1.3. Frequencies and Updating

As indicated by the discussion of (1.2), one formal contribution of the paper is to generalize de Finetti's Theorem from probability measures to belief functions. However, the importance of the de Finetti Theorem extends beyond the representation to the connection it affords between subjective beliefs and empirical frequencies. Here we outline how these aspects of the de Finetti model extend also to our generalization.

One form that the noted connection takes in the Bayesian framework is to relate subjective beliefs about the unknown but fixed bias (or more general parameter), represented by $\mu$, to empirical frequencies. In the coin-tossing setting, for example, empirical frequencies converge with probability 1 (by a law of large numbers for exchangeable measures), and one can view $\mu$ as representing ex ante beliefs about the limiting empirical frequency of Heads, a random variable. Thus a bridge is provided between subjectivist and frequentist theories of probability (see Kreps [25, Ch. 11], for example). Secondly, this connection can help a decision-maker to calibrate her uncertainty about the true parameter. Another important aspect of the de Finetti Theorem is the connection between beliefs and observations afforded via Bayesian updating of the prior $\mu$. 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 [36, Ch. 3.6], for example).

Our generalization of de Finetti's Theorem also extends the scope of these contributions. The link to a frequentist foundation for beliefs extends in the following way (for the coin-tossing experiments): Given ambiguity, the decision-maker is not certain that empirical frequencies converge to a fixed point. Thus she thinks in terms of the random interval $[\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)]$, where $\Psi_n(\omega)$ is the empirical frequency of Heads for a sample of size $n$. We show (Section 5), using a LLN for belief functions due to Maccheroni and Marinacci [27], that the prior $\mu$ over $Bel(S)$ can be viewed as a measure over such random intervals. An aid to forming beliefs about belief functions is also provided thereby.

Turn to updating. Ambiguity poses difficulties for updating and there is no consensus updating rule analogous to Bayes' Rule. However, our 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, our model of updating is also tractable. This aspect stems from the fact that beliefs at every node have a representation of the form (1.2), which is completely defined by a (unique) probability measure over $Bel(S)$. 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 or a nonadditive measure or capacity. 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 6.4) a counterpart of the Savage result that (under suitable conditions), 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.

## 2. BELIEF FUNCTIONS

We will deal with two different (compact metric) state spaces - $S$ corresponding to a single experiment, and $\Omega = S^{\infty}$, describing all possible sample paths. Thus in this section we consider an abstract (compact metric) state space $X$. It has Borel $\sigma$-algebra $\Sigma_X$.

A *belief function* on $X$ is a set function $\nu : \Sigma_X \to [0,1]$ such that:[5]

Bel.1 $\nu(\varnothing) = 0$ and $\nu(X) = 1$

Bel.2 $\nu(A) \subset \nu(B)$ for all Borel sets $A \subset B$

Bel.3 $\nu(B_n) \downarrow \nu(B)$ for all sequences of Borel sets $B_n \downarrow B$

Bel.4 $\nu(G) = \sup\{\nu(K) : K \subset G,\ K \text{ compact}\}$, for all open $G$

Bel.5 $\nu$ is totally monotone (or $\infty$-monotone): for all Borel sets $B_1, .., B_n$,

$$\nu\left(\cup_{j=1}^{n} B_j\right) \geq \sum_{\varnothing \neq J \subset \{1,\ldots,n\}} (-1)^{|J|+1} \nu\left(\cap_{j \in J} B_j\right)$$

The set of all belief functions on $X$ is $Bel(X)$. It is compact metric when endowed with the topology for which $\nu_n \to \nu$ if and only if $\int f d\nu_n \to \int f d\nu$ for every continuous function $f$ on $X$, where the integral here and throughout is in the sense of Choquet (see Schmeidler [37]).

Denote by $\Delta(X)$ the set of Borel probability measures on $X$, endowed with the weak convergence topology (generated by continuous functions), and by $\mathcal{K}(X)$ the set of compact subsets of $X$, endowed with the Hausdorff metric. Both are

[5] These conditions are adapted from [35], to which we refer the reader for details supporting much of the outline in this section. We point out only that when restricted to probability measures, Bel.4 is the well-known property of regularity.

compact metric. If $m \in \Delta(X)$, then $m^\infty$ denotes the usual i.i.d. product measure on $X^\infty$.[6]

Each belief function defines a preference order or utility function. Interpreting $X$ as a state space, denote by $\mathcal{F}(X)$ the set of all (measurable) acts $f : X \to [0,1]$. For any $\nu \in Bel(X)$, let $U_\nu : \mathcal{F}(X) \to \mathbb{R}$ be defined by

$$U_\nu(f) = \int f d\nu. \tag{2.1}$$

Refer to $U_\nu$ as a *belief-function utility*, and to the corresponding preference order as a *belief-function preference*.

Since a belief function is also a capacity (defined by Bel.1-Bel.4), belief-function utility is a special case of Choquet expected utility, axiomatized by Schmeidler [37]; and since it is convex (supermodular, or 2-alternating, that is, satisfies the inequalities in Bel.5 restricted to $n = 2$), it is well-known that (2.1) can be expressed alternatively as

$$U_\nu(f) = \min_{P \in core(\nu)} \int f dP,$$

where

$$core(\nu) = \{P \in \Delta(X) : P(\cdot) \geq \nu(\cdot)\}.$$

Thus the current model is also a special case of multiple-priors utility, which has been axiomatized by Gilboa and Schmeidler [20].

Note that acts are taken to be real-valued and to enter linearly into the Choquet integral. This may be justified as follows: Suppose that outcomes lie in an abstract set $Z$, and that (Anscombe-Aumann) acts map states into $\Delta(Z)$. Suppose also that there exist best and worst outcomes $\overline{z}$ and $\underline{z}$. Then, under weak conditions, for each state $\omega$ 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*.[7] 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 $U_\nu(f)$ is also scaled in probability units - it satisfies

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

[6] Throughout product spaces are endowed with the product metric.

[7] 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.

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

Belief functions have been widely studied (see [8, 30, 33], for example) and used (for applications in robust statistics see Huber [24], and for applications in decision theory and economics, see [15, 35, 27], for example). They (and their corresponding utility functions) also admit an intuitive justification due to Dempster [9] and Shafer [38], (see also Mukerji [32] and Ghirardato [18]), and illustrated in (2.3) below.

Though there exists a Savage-style state space $X$, the agent's perceptions are coarse and are modeled through an auxiliary epistemic (compact metric) state space $\widehat{X}$ and a (measurable and nonempty-compact-valued) correspondence $\Gamma$ from $\widehat{X}$ into $X$. There is a Borel probability measure $p$ representing beliefs on $\widehat{X}$.

$$\begin{array}{ccc} (\widehat{X}, p) & \overset{\Gamma}{\rightsquigarrow} & (X, \nu) \\ & \searrow & \downarrow_f \\ & & [0,1] \end{array} \tag{2.3}$$

A Bayesian agent would view each physical action as an act from $X$ to the outcome set $[0,1]$, and would evaluate it via its expected utility (using a probability measure on $X$). The present agent is aware that while she can assign probabilities on $\widehat{X}$, events there are only imperfect indicators of payoff-relevant events in $X$. Such awareness and a conservative attitude lead to preference that can be represented by the utility function

$$U^{DS}(f) = \int_{\widehat{X}} \left( \inf_{x \in \Gamma(\widehat{x})} f(x) \right) dp = \int_X f d\nu,$$

where $\nu$ is the belief function given by[8]

$$\nu(A) = p\left(\left\{\widehat{x} \in \widehat{X} : \Gamma(\widehat{x}) \subset A\right\}\right), \text{ for every } A \in \Sigma_X.$$

As a foundation for belief function utility, the preceding is suggestive though limited, because $\widehat{X}$, $p$ and $\Gamma$ are presumably not directly observable. However, Epstein, Marinacci and Seo [10, Section 4.3] describe behavioral foundations for a Dempster-Shafer representation.

[8] We can view $\Gamma$ as a function from $\widehat{X}$ to $\mathcal{K}(X)$. Then $\Gamma$ is measurable [3, Thm. 18.10] and induces the measure $p \circ \Gamma^{-1}$ on $\mathcal{K}(X)$. Choquet's theorem (see below) implies that $\nu(\cdot) = p \circ \Gamma^{-1}(\{K : K \subset \cdot\})$ is a belief function.

A central fact about belief functions is the Choquet Theorem [35, Thm. 2].[9] To state it, note that, by [35, Lemma 1], $\{K \in \mathcal{K}(X) : K \subset A\}$ is universally measurable for every $A \in \Sigma_X$. Further, any Borel probability measure (such as $m$ on $\Sigma_{\mathcal{K}(X)}$) admits a unique extension (also denoted $m$) to the collection of all universally measurable sets.[10]

**Theorem 2.1 (Choquet).** *The set function $\nu : \Sigma_X \to [0,1]$ is a belief function if and only if there exists a (necessarily unique) Borel probability measure $m_\nu$ on $\mathcal{K}(X)$ such that*

$$\nu(A) = m_\nu(\{K \in \mathcal{K}(X) : K \subset A\}), \text{ for every } A \in \Sigma_X. \tag{2.4}$$

*Moreover, in that case, for every act $f$,*

$$U_\nu(f) = \int_X f d\nu = \int_{\mathcal{K}(X)} \left( \inf_{x \in K} f(x) \right) dm_\nu. \tag{2.5}$$

The one-to-one mapping $\nu \longmapsto m_\nu$ is denoted $\zeta$. It constitutes a homeomorphism between $Bel(X)$ and $\Delta(\mathcal{K}(X))$. One perspective on the theorem is that it shows that any belief function has a special Dempster-Shafer representation where

$$\widehat{X} = \mathcal{K}(X), \Gamma(K) = K \subset X \text{ and } p = m_\nu.$$

Conclude this overview of belief functions with a simple example. Let $X = \{H, T\}$ and $[p_m, p_M] \subset [0,1]$, thought of as an interval of probabilities for Heads. Define $\theta$ on subsets of $X$, by $\theta(H) = p_m$, $\theta(T) = 1 - p_M$, and $\theta(X) = 1$. Then $\theta$ is a belief function - the measure $m$ from Choquet's Theorem is $m(H) = \theta(H)$, $m(T) = \theta(T)$ and $m(\{H,T\}) = 1 - \theta(T) - \theta(H)$, the length of the interval. An interpretation is that the coin is seen as being drawn from an urn containing many coins, of which the proportion $\theta(H)$ ($\theta(T)$) are sure to yield Heads (Tails), and where there is complete ignorance about the remaining proportion. In particular, for binary state spaces, a belief function can be thought of simply as a probability interval.

[9] The final assertion in the theorem stated below relies also on [35, Thm. 3]. See also [30, Thm. 5.1] and [7, Thm. 3.2].

[10] Throughout, given any Borel probability measure, we identify it with its unique extension to the $\sigma$-algebra of universally measurable sets.

## 3. THE MODEL

While the above refers to an abstract state space $X$, we assume a specific structure here corresponding to the presence of many experiments. Thus let $S$ be a compact metric space thought of as the stage state space, or the set of possible outcomes for any single experiment. The full state space is

$$\Omega = S_1 \times S_2 \times \; .... = S^{\infty}, \text{ where } S_i = S \text{ for all } i.$$

Denote by $\Sigma_i$ the Borel $\sigma$-algebra on $S_i$, which can be identified with a $\sigma$-algebra on $\Omega$, and by $\Sigma_\Omega$, the product Borel $\sigma$-algebra. For the most part, we take the abstract space $X$ above to be $\Omega$, and in that case, when there is no danger of confusion, we abbreviate $\Sigma_\Omega$, $\mathcal{K}(\Omega)$, $Bel(\Omega)$ and $\mathcal{F}(\Omega)$ by $\Sigma$, $\mathcal{K}$, $Bel$ and $\mathcal{F}$. For any $I \subset \mathbb{N}$, $\Sigma_I$ is the $\sigma$-algebra generated by $\{\Sigma_i : i \in I\}$ and $\mathcal{F}_I$ denotes the set of $\Sigma_I$-measurable acts. An act is said to be *finitely-based* if it lies in $\cup_{I \text{ finite}} \mathcal{F}_I$.

We will have occasion to refer also to $\mathcal{K}(S)$ and $Bel(S)$.

We are given a belief-function preference $\succeq$ and the corresponding utility function $U$. We adopt two axioms for preference (or equivalently, for utility) that describe the individual's perception of how experiments are related.

Let $\Pi$ be the set of (finite) permutations on $\mathbb{N}$. For $\pi \in \Pi$ and $\omega = (s_1, s_2, ...) \in \Omega$, let $\pi\omega = \left(s_{\pi(1)}, s_{\pi(2)}, ...\right)$. Given an act $f$, define the permuted act $\pi f$ by $(\pi f)(s_1, ..., s_n, ...) = f\left(s_{\pi(1)}, ..., s_{\pi(n)}, ...\right)$.

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

$$f \sim \pi f.$$

Our second axiom relaxes the Independence axiom by permitting value for randomization (or mixing) as a way to alleviate the concern that experiments may differ. First, working within the coin-tossing example, we illustrate why randomization can be valuable.

Symmetry implies the indifference

$$H_1 T_2 \sim T_1 H_2.$$

Here $H_1 T_2$ is the bet that pays 1 util if the first toss yields Heads and the second Tails; the bet $T_1 H_2$ 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 she admits the *possibility* that technique varies, and hence that experiments are not identical, then the individual 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" states 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

$$\tfrac{1}{2}H_1T_2 + \tfrac{1}{2}T_1H_2 \succ H_1T_2 \sim T_1H_2, \tag{3.1}$$

contrary to the Independence Axiom.

On the other hand, in well-defined cases, mixing is a matter of indifference just as in expected utility theory. We capture behaviorally the perception of poorly understood and idiosyncratic factors influencing experiments by specifying precisely when mixing can conceivably be valuable; we do this by specifying when it definitely cannot be valuable.

One might expect that randomization will not matter when mixing bets on different experiments. For example, we would expect the ranking

$$\tfrac{1}{2}H_1 + \tfrac{1}{2}H_2 \sim H_1 \sim H_2.$$

The reason is that, even though the payoff to $H_1$ is uncertain, $H_2$ does not hedge this uncertainty if tossing technique is perceived to be unrelated across tosses. Neither does it hedge uncertainty about the physical bias - since the bias is identical for both tosses, mixing $H_1$ and $H_2$ does not moderate payoff uncertainty. Thus the mixture $\frac{1}{2}H_1 + \frac{1}{2}H_2$ is no better than either component bet.

As another example, suppose that $H_2 \succeq T_2$ and ask whether this ranking should remain invariant after mixing each bet with $H_3$, that is, consider

$$\tfrac{1}{2}H_2 + \tfrac{1}{2}H_3 \succeq \tfrac{1}{2}T_2 + \tfrac{1}{2}H_3.$$

There is intuition *against* this ranking if (and only if) there is ambiguity about the physical bias, because then mixing $H_3$ and $T_2$ smooths out ambiguity about the bias, while the same is not true for $\frac{1}{2}H_2 + \frac{1}{2}H_3$, and thus the individual may strictly prefer $\frac{1}{2}T_2 + \frac{1}{2}H_3$.

With these illustrations in mind, we can now state our central axiom, which relaxes the Independence axiom. It expresses both the noted stochastic independence across experiments and the absence of ambiguity about the factors (such as the bias of the single coin) that are common to all experiments.[11]

**Axiom 2 (ORTHOGONAL INDEPENDENCE (OI)).** *For all* $0 < \alpha \leq 1$, *and acts* $f', f \in \mathcal{F}_I$ *and* $g \in \mathcal{F}_J$, *with* $I$ *and* $J$ *finite and disjoint,*

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

It is easy to show that OI is satisfied if and only if $U$ satisfies: For all $f \in \mathcal{F}_I$ and $g \in \mathcal{F}_J$, where $I$ and $J$ are finite and disjoint, and for all $\alpha$ in $[0,1]$,

$$U(\alpha f + (1-\alpha) g) = \alpha U(f) + (1-\alpha) U(g). \tag{3.2}$$

We use this characterization of OI frequently.

## 4. THE REPRESENTATION

Given $\theta \in Bel(S)$, we define an "i.i.d. product" $\theta^\infty$, a belief function on $S^\infty$, as follows. We have $\zeta(\theta) \in \Delta(\mathcal{K}(S))$, and hence $(\zeta(\theta))^\infty \in \Delta[(\mathcal{K}(S))^\infty] \subset \Delta[(\mathcal{K}(S^\infty))]$.[12] By the Choquet Theorem, there exists a (unique) belief function on $S^\infty$ corresponding to $(\zeta(\theta))^\infty$. Denote it by $\theta^\infty$, so that

$$\zeta(\theta^\infty) = (\zeta(\theta))^\infty.$$

Since Hendon et al [22] propose this rule in the case of finite Cartesian products, we refer to $\theta^\infty$ as the *Hendon i.i.d. product*. Ghirardato [19, Theorem 3] shows that it is the only product rule for belief functions such that the product (i) is also a belief function, and (ii) it satisfies a mathematical property called the Fubini property. The Hendon product is central also in our model, but here it will emerge as an implication of assumptions about preference. Finally, note that when $\theta$ is additive, and thus a probability measure, then $\theta^\infty$ is the usual i.i.d. product.

[11] We do not have a more general model where ambiguity about common factors is permitted. See the concluding section for further discussion.

[12] Recall that $\zeta$ denotes the homeomorphism defined by Theorem 2.1. We denote by $\nu$ a generic belief function on $S^\infty$ and by $\theta$ a generic belief function on $S$.

The belief-function utility $V$ on $\mathcal{F}$ is called an *i.i.d. (belief-function) utility* if there exists $\theta \in Bel\left(S\right)$ such that

$$V\left(f\right) = V_{\theta^{\infty}}\left(f\right) \equiv \int f d\left(\theta^{\infty}\right), \text{ for all } f \in \mathcal{F}.$$

Our main result is that Symmetry and OI characterize utility functions that are "mixtures" of i.i.d. utilities.

Before stating the theorem, we describe a sense in which each i.i.d. utility function $V_{\theta^{\infty}}$ captures (stochastic) independence across experiments. It is easily verified (using (2.4)) that[13]

$$\theta^{\infty}\left(A_I \times A_J \times S^{\infty}\right) = \theta^{\infty}\left(A_I \times S^{\infty}\right)\ \theta^{\infty}\left(A_J \times S^{\infty}\right), \tag{4.1}$$

for $A_I \in \Sigma_I$ and $A_J \in \Sigma_J$, where $I, J \subset \mathbb{N}$ are finite and disjoint. Though the interpretation of (4.1) as expressing stochastic independence may seem obvious, we elaborate on this interpretation. Identify each event $E \subset \Omega$ with the corresponding bet, the act defined by the indicator function $\mathbf{1}_E$, and note that $V_{\theta^{\infty}}\left(\mathbf{1}_E\right) = \nu\left(E\right)$. Recall also that the outcomes of acts are measured in probability units, as are utilities (see (2.2)). Therefore, $\mathbf{1}_E$ is indifferent to betting on the toss of an objective $\nu\left(E\right)$-coin. This is true in particular for each of the three events appearing in (4.1).[14] Thus, a bet on $A_I$ is indifferent to a bet on an objective $\nu\left(A_I\right)$-coin, a bet on $A_J$ is indifferent to a bet on an objective $\nu\left(A_J\right)$-coin, and a bet on $A_I \times A_J$ is indifferent to a bet on the toss of an objective coin where the probability of winning is $\nu\left(A_I\right)\nu\left(A_J\right)$, that is, to the bet on successive wins in *independent* tosses of the $\nu\left(A_I\right)$- and $\nu\left(A_J\right)$-coins.

The latter independence of the objective coins is strongly suggestive of the asserted perception of stochastic independence of experiments. Alternative support for the latter interpretation is given below by our model of updating, according to which an i.i.d. utility function (or belief function) is never modified in response to past observations.

We can finally state our representation result.

**Theorem 4.1.** *Let $U$ be a belief function utility. Then the following statements are equivalent:*

*(i) $U$ satisfies Symmetry and Orthogonal Independence.*

[13] A more general "product relation" involving (nonindicator) acts is aso satisfied (see [11]).

[14] Identify $A_I \times S^{\infty}$ with $A_I$ and so on.

*(ii) There exists a (necessarily unique) Borel probability measure $\mu$ on $Bel(S)$ such that*

$$U(f) = \int_{Bel(S)} V_{\theta^{\infty}}(f)\, d\mu(\theta), \text{ for every } f \text{ in } \mathcal{F}. \tag{4.2}$$

*(iii) There exists a (necessarily unique) Borel probability measure $\mu$ on $Bel(S)$ such that $\nu$, the belief-function corresponding to $U$, can be expressed in the form*

$$\nu(A) = \int_{Bel(S)} \theta^{\infty}(A)\, d\mu(\theta), \text{ for every } A \text{ in } \Sigma. \tag{4.3}$$

As emphasized earlier, we interpret the de Finetti Theorem as a result regarding preference that assumes subjective expected utility. With this interpretation, we generalize his result to the framework of belief function preference.

The more general representation (4.3) also admits a "conditionally i.i.d." interpretation. However, as explained in the introduction (following (1.2)), the fact that each $\theta$ is a belief function rather than a probability measure, accommodates the perception that experiments are indistinguishable but not necessarily identical. Speaking informally, indistinguishability is delivered because the same $\theta$ applies to each experiment, while "not identical" is captured essentially because any probability measure in the core of $\theta$ could apply to any experiment.[15]

Another contribution of the theorem is as an aid to a decision-maker in forming beliefs. Paralleling de Finetti's contribution for a Bayesian framework, the representation (1.2) can aid in forming a (probabilistic) prior even where experiments are ambiguous. For example, it is arguably easier to decide on which intervals might describe every coin and on a probability distribution over them than to arrive at beliefs, in the form of a belief function, directly over all possible sample realizations.

It is evident that the axiomatic characterization of (4.2) provided by the theorem is tight - both axioms are necessary and neither is implied by the other. An expected utility function with prior $p$ on $\Omega$ that is not exchangeable, satisfies OI but not Symmetry. For a 'dual' example, let $K^{*} = \{(H, H, ...), (T, T, ...)\}$ and

$$\nu(A) = \begin{cases} 1 & A \supset K^{*} \\ 0 & \text{otherwise.} \end{cases}$$

[15] This generalizes the informal description in terms of probability intervals given in the introduction for the case of a binary state space $S = \{H, T\}$.

Then, by the Choquet Theorem, $\nu$ is a belief function on $\Omega$. The corresponding utility function, as in (2.1), satisfies Symmetry but not OI.[16] For example, it implies

$$U\left(\tfrac{1}{2}H_1 + \tfrac{1}{2}T_2\right) = \tfrac{1}{2} > 0 = \tfrac{1}{2}U\left(H_1\right) + \tfrac{1}{2}U\left(T_2\right),$$

contradicting (3.2). The reason is that according to $\nu$, coin tosses are perceived as perfectly correlated and thus decidedly related to one another.

Finally, for those readers who are concerned about practicality, we offer some reassurance. One obstacle to evaluating the utility functions appearing in the theorem, is evaluation of i.i.d. utility functions $V_{\theta^\infty}$, and thus expressions of the form

$$V_{\theta^\infty}(f) \equiv \int f d\left(\theta^\infty\right).$$

But the Choquet Theorem and the definition of $\theta^\infty$ make this straightforward. Let $m_\theta \in \Delta\left(\mathcal{K}(S)\right)$ denote the measure on subsets implied by Theorem 2.1, in which case the i.i.d. product $(m_\theta)^\infty \in \Delta\left[\left(\mathcal{K}(S)\right)^\infty\right]$ is the corresponding measure for $\theta^\infty$. Then, by (2.5),

$$V_{\theta^\infty}(f) = \int_{(\mathcal{K}(S))^\infty} \left( \inf_{\omega \in K_1 \times K_2 \times \ldots} f(\omega) \right) \ d(m_\theta)^\infty \left(K_1, K_2, \ldots\right),$$

which is a standard integral with respect to an additive product measure. One need only derive $m_\theta$ for each relevant $\theta$. For a binary state space, the one-to-one map between $\theta$ and $m_\theta$ was described at the end of Section 2. Also more generally, if $S$ is finite, then $m_\theta$ can be constructed explicitly from $\theta$ by the so-called Mobius inversion formula[17]

$$m_\theta(A) = \Sigma_{B \subset A} (-1)^{\#(A \backslash B)} \theta(B), \text{ if } A \subset S.$$

See Hendon *et al* [22, p. 100], for example.

Another possible concern is whether a decision-maker can plausibly arrive at a prior $\mu$ over belief functions. However, belief functions are often not such complicated objects. For example, with repeated coin-tossing, when $S$ is binary, each belief function $\theta$ corresponds to a unique interval, and the latter corresponds to an ordered pair of real numbers - in other words, one need only formulate a prior over

[16] Talagrand [40] contains the study of symmetric belief functions, where OI for the corresponding utility function is not assumed.

[17] Since $m_\theta$ is additive, it is uniquely defined as a measure on $\mathcal{K}(S)$ by its values $m_\theta(A)$ for every $A \subset S$.

an unknown two-dimensional parameter. More generally, since each $\theta$ corresponds to a unique Mobius inverse $m_{\theta}$, the task is to form a prior over $\Delta(\mathcal{K}(S))$. This is perhaps more difficult than forming a prior over $\Delta(S)$, as required by de Finetti, but is qualitatively comparable to the latter. Some guidance for our agents in calibrating beliefs is provided next, when the connection to empirical frequencies is considered.

## 5. FREQUENCIES

In this section, we relate subjective uncertainty about the true i.i.d. belief function $\theta^{\infty}$, represented by $\mu$, to beliefs about empirical frequencies. Formally, our result is a corollary of our de Finetti-style representation and a law of large numbers (LLN) for belief functions due to Maccheroni and Marinacci [27].

The coin-tossing setting conveys the point most clearly. Let $S = \{H, T\}$ and denote by $\Psi_n(\omega)$ the proportion of Heads realized in the first $n$ experiments along the sequence $\omega$. Then, for any $\theta \in Bel(S)$, the noted LLN for belief functions implies that

$$\theta^{\infty}\left(\left\{\theta(H) \leq \liminf \Psi_n(\omega) \leq \limsup \Psi_n(\omega) \leq 1 - \theta(T)\right\}\right) = 1. \tag{5.1}$$

Further, these bounds on empirical frequencies are tight in the sense that[18]

$$\begin{aligned} [a &> \theta(H) \text{ or } b < 1 - \theta(T)] \implies 0 = \\ &\theta^{\infty}\left(\left\{a \leq \liminf \Psi_n(\omega) \leq \limsup \Psi_n(\omega) \leq b\right\}\right). \end{aligned} \tag{5.2}$$

Therefore, the representation (4.3) implies that, for every $0 \leq a \leq b \leq 1$,

$$\begin{aligned} &\mu\left(\left\{\theta : a \leq \theta(H) \leq 1 - \theta(T) \leq b\right\}\right) \\ = \ &\nu\left(\left\{a \leq \liminf \Psi_n(\omega) \leq \limsup \Psi_n(\omega) \leq b\right\}\right). \end{aligned} \tag{5.3}$$

This equality admits an appealing interpretation. In the Bayesian setting, each coin toss is described by a common unknown probability of Heads, and the LLN justifies interpreting uncertainty about this "parameter" in terms of uncertainty about the limiting empirical frequency of Heads. In our setting, the individual is not certain that empirical frequencies converge to a fixed point, and she thinks in

[18] This is proven in the context of proving the Corollary below.

terms of intervals that will contain all limit points. Supposing for simplicity that $\mu$ has finite support, then (5.3) is equivalent to:

$$\mu(\theta) = \nu(\{\omega : \theta(H) \leq \liminf \Psi_n(\omega) \leq \limsup \Psi_n(\omega) \leq 1 - \theta(T)\}). \qquad (5.4)$$

Thus the prior subjective probability of the unknown (but nonrandom) parameter $\theta$ equals the prior likelihood, according to $\nu$, that the interval $[\theta(H), 1-\theta(T)]$ will contain the random interval of empirical frequencies in the long run. This provides a frequentist perspective for the probability measure $\mu$ over belief functions.

Consistent with the normative slant of our model, it is also worthwhile noting that (5.4) can also help a decision-maker, a statistician for example, to calibrate her uncertainty about the true $\theta$. That is because $\mu(\theta)$ equals that prize which, if received with certainty, would be indifferent for her to betting (with prizes 1 and 0) on the event that

$$[\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset [\theta(H), 1 - \theta(T)].$$

We elaborate briefly on the formal meaning of the preceding. Any $\theta \in Bel(S)$ is completely determined by the two numbers $\theta(H)$ and $\theta(T)$, or equivalently by the interval

$$I_\theta = [\theta(H), 1 - \theta(T)].$$

Moreover, $\theta \longmapsto I_\theta$ is one-to-one. Thus the representing measure $\mu$ can be thought of as a measure over intervals $I_\theta$. Formally, let $\mathcal{I}$ be the collection of all compact subintervals of $[0,1]$. As a subset of $\mathcal{K}([0,1])$, $\mathcal{I}$ inherits the Hausdorff metric and the associated Borel $\sigma$-algebra. Moreover, $Bel(S)$ is homeomorphic to $\mathcal{I}$, and thus there is a one-to-one correspondence, denoted $e$, between probability measures on $Bel(S)$, and probability measures on intervals, that is, measures in $\Delta(\mathcal{I})$. In particular, $\mu$ can be identified with a unique $\widehat{\mu} = e(\mu)$ in $\Delta(\mathcal{I})$. Thus (5.3) can be written in the form

$$\begin{aligned} & \widehat{\mu}(\{I : I \subset [a,b]\}) \\ = \; & \nu(\{\omega : [\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset [a,b]\}). \end{aligned} \qquad (5.5)$$

The general (nonbinary) case is similar. Denote by $\Psi_n(\cdot)(\omega)$ the empirical frequency measure given the sample $\omega$; $\Psi_n(B)(\omega)$ is the empirical frequency of $B \in \Sigma_S$ in the first $n$ experiments. The above reasoning can be extended to prove:

**Corollary 5.1.** *Let $U = U_\nu$ be a belief function utility. Then the equivalent statements in Theorem 4.1 are equivalent also to the following: There exists a* *probability measure $\mu$ on $Bel\,(S)$ satisfying both (i) $\mu$ represents $U$ in the sense of (4.2); and (ii) for every finite collection $\{A_1, ..., A_I\} \subset \Sigma_S$, and for all $a_i \leq b_i$, $i = 1, ..., I$,*

$$\begin{aligned} & \mu\left(\bigcap_{i=1}^{I} \{\theta : [\theta\,(A_i)\,, 1 - \theta\,(S \backslash A_i)] \subset [a_i, b_i]\}\right) && (5.6) \\ = \;\; & \nu\left(\bigcap_{i=1}^{I} \{\omega : [\liminf \Psi_n\,(A_i)\,(\omega)\,, \limsup \Psi_n\,(A_i)\,(\omega)] \subset [a_i, b_i]\}\right). \end{aligned}$$

Equation (5.6) relates the prior $\mu$ over belief functions to ex ante beliefs about empirical frequencies for the events $A_1, ..., A_I$. More precisely, the $\mu$-measures of only the sets shown are so related. However, as our final result shows, $\mu$ is completely determined by its values on these sets.

**Proposition 5.2.** *If $\mu, \mu' \in \Delta\,(Bel\,(S))$ coincide on all sets of the form*

$$\{\theta \in Bel\,(S) : \theta\,(A_1) \geq a_1, ..., \theta\,(A_I) \geq a_I\}\,,$$

*where $A_i, a_i$ and $I$ vary over $\Sigma_S$, $[0,1]$ and the positive integers respectively, then $\mu = \mu'$.*

# 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 $n$. Acts over the set $I \subset \mathbb{N}$ of experiments lie in $\mathcal{F}_I$; abbreviate $\mathcal{F}_{\{i\}}, \mathcal{F}_{\{1,...n\}}$ and $\mathcal{F}_{\{n+1,n+2,...\}}$ by $\mathcal{F}_i$, $\mathcal{F}_{\leq n}$ and $\mathcal{F}_{>n}$. In this section, we assume that $S$ is finite.

Ex ante preference on $\mathcal{F}$ is $\succeq$ , 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 $\succeq$ 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.

We assume 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.

*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 $(s_{n+1}, s_{n+2}, ...)$.

A final assumption is that $\succeq$ and every $\succeq_n$ satisfy the axioms of our belief function model.[19] Call this composite axiom *Exchangeability.*

[19] In fact, the main result below (Theorem 6.1) relies on the existence of a de Finetti-style representation, but not on belief functions per se. Thus it can be translated into a model of updating for our multiple-priors-based analysis of exchangeability in [11]. We discuss that model below in the concluding section.

Since all utility functions satisfy our axioms, each admits a representation in terms of a unique measure over $Bel(S)$, the set of belief functions over $S$. Their utility functions are $U$ and $U_n(\cdot \mid s_1^n)$, for $\succeq$ and $\succeq_{n,s_1^n}$ respectively; frequently, dependence on the sample is suppressed and we write simply $\succeq_n$ and $U_n$. Then

$$U(f) = \int_{Bel(S)} V_{\theta^\infty}(f)\, d\mu(\theta), \text{ for all } f \in \mathcal{F},$$

where $\theta^\infty$ is the Hendon product and $V_{\theta^\infty}$ is the corresponding IID utility function. Similarly, imposing Consequentialism,

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

for some probability measure $\mu_n$ that depends on the realized sample $s_1^n$. The updating problem thus reduces to describing the evolution of $\mu_n$ as a function of $\mu_0 = \mu$ and the realized sample.

The implications of WDC and the other axioms are described in terms of a *likelihood function* $L : Bel(S) \to \Delta(\Omega)$, where $\theta \longmapsto L(B \mid \theta)$ is (Borel) measurable for each measurable subset $B$ of $\Omega$. Think of $L(B \mid \theta)$ as the likelihood of $B \subset \Omega$, a set of infinite samples, conditional on $\theta$ describing each experiment. 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 $n$, $L_n : S^{n-1} \times Bel(S) \to \Delta(S)$.[20] Thus for each sample $s_1^{n-1}$, $L_n(\cdot \mid \theta) \in \Delta(S)$ gives the probability distribution, or likelihood, for the $n^{th}$ experiment, conditional on $s_1^{n-1}$ and for the given $\theta$.

The central result in our model of updating follows.

**Theorem 6.1.** *The axioms Consequentialism, WDC and Exchangeability 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 $n \geq 1$,*

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

*where*

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

*is a probability measure on $S$ having full support.*

[20] More precisely, $L_n(\cdot \mid \theta)$ 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 $f \in \mathcal{F}_{>n}$,

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

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 C 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 [29]); other special cases include Harsanyi's aggregation theorem [21] and probability aggregation [31].

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. The broader tenet implicit in our model is that prior ambiguity about signals is *conceptually separate* from (albeit related to) inference after realization of the signal. Prior ambiguity about signals is primarily about how to rank bets on the signal ex ante, while the inference problem concerns the ex post interpretation of the signal as information about other experiments.

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

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

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.[21] 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. Alternatively, note that $\overline{L}(\cdot)$ and its conditionals describe beliefs about future payoff relevant states in the Bayesian context, while in our model "beliefs" at node $n$ are represented by the belief function $\int \theta^{\infty} d\mu_n$.) 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 preferences being represented by additive measures $\mu_n$. 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 non-identical.

Before examining further axioms, it is convenient to clarify first uniqueness properties. Define the process $\{w_n\}$ by

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

Refer to $w_n(s_n,\theta)$ as the *weight of evidence* for $\theta$ 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

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

all $n$. Then any specification with $L_n(\cdot;\theta) = \overline{L}_n(\cdot)$, 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 $L_n(\cdot;\theta) = w_n(\cdot;\theta)\,\overline{L}_n(\cdot)$ are also unique for each $s_n$. Uniqueness of $L$ follows (up to $\mu$-nullity).

We summarize the preceding more formally. First, we add the axiom:[22]

*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 $n$,*

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

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

### 6.2. Further Restrictions

Weak Dynamic Consistency alone does not pin down the process of posterior beliefs and preferences emanating from a given ex ante preference - any likelihood function $L$ leads, via (6.1), to satisfaction of WDC. Here we explore two further axioms and the restrictions they impose on $L$. Such restrictions are meaningful given Non-Collinearity, by the uniqueness established in Corollary 6.2. Since it is simplifying, we adopt Non-Collinearity below. However, it should be possible to use the weights process to describe the implications of the axioms more generally.

Above we emphasized the distinction between ex ante ambiguity about a signal and ex post inference from realization of the signal. The next axiom imposes a connection.

*Payoff Ambiguity*: For every $n$, $f \in \mathcal{F}_{>n}$ and $f_n \in \mathcal{F}_{\leq n}$, if $f \sim_{n,s_1^n} f_n$ for all $s_1^n$, then $f \succeq f_n$.

Suppose that for every sample $s_1^n$, after observing it and updating beliefs accordingly, the individual is indifferent between $f$, which depends only on future

[22] 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.

experiments, and $f_n(s_1^n)$, a deterministic outcome since $f_n \in \mathcal{F}_{\leq n}$. In other words, for any $s_1^n$, $f_n(s_1^n)$ is the ex post "value" of holding $f$ conditional on $s_1^n$. Dynamic consistency, in the form of DC, would imply indifference between $f$ and $f_n$ also ex ante.[23] However, we have indicated a difference between the individual's ex ante and ex post perspectives that argues against DC. When evaluating $f_n$, the sample $s_1^n$ is payoff relevant and ex ante ambiguity about the first $n$ experiments matters. By contrast, such ex ante ambiguity is not relevant to the evaluation of $f$ since for it $s_1, ..., s_n$ serve as pure signals. In this case, ex ante ambiguity about the first $n$ experiments may translate into a concern with the quality of the information provided by the sample. The ex post concern with quality is already incorporated into the conditional indifference $f \sim_{n,s_1^n} f_n$. If the assessment (or anticipation) of quality is identical also ex ante, then the only difference between the two stages is the added ambiguity associated with $f_n$ when viewed from the ex ante perspective. Thus, in going back to the ex ante stage, $f_n$ is discounted relative to $f$, which is thus weakly preferable.

**Corollary 6.3.** *Let $\succeq$ and $\{\succeq_n\}$ satisfy Consequentialism, WDC, Exchangeability and Non-Collinearity. Let $L$ be the (essentially unique) likelihood function provided by Theorem 6.1 and Corollary 6.2, and let $\overline{L}$ be defined by (6.4). Then Payoff Ambiguity is satisfied if and only if*

$$\overline{L}(\cdot) \in core\left(\int \theta^\infty d\mu\right). \tag{6.6}$$

Payoff Ambiguity restricts the prior of the shadow Bayesian model to lie in the core of $\nu = \int \theta^\infty d\mu$, the belief function on $\Omega$ that defines ex ante preference (see Theorem 4.1). As a result, a connection is imposed between ex post inference and ex ante preference.

Condition (6.6) is sufficient even without Non-Collinearity.[24] It has the following further implication. Since posteriors form an $\overline{L}$-martingale, for any $\theta$, $\mu_\infty(\theta)$ exists except for an $\overline{L}$-null set $B$. But, by (6.6),

$$\overline{L}(B) = 0 \Longrightarrow U(\mathbf{1}_B) = 0, \tag{6.7}$$

in other words, the individual would be unwilling to bet against posteriors at $\theta$ converging. This is a limited sense of certainty that beliefs will converge.

[23] WDC does not cover this situation since $f_n$ depends on the $n$th experiment, while WDC applies only to the ranking of acts that are independent of signals.

[24] The stronger condition $L(\cdot \mid \theta) \in core(\theta^\infty)$ $\mu$-a.s. seems particularly natural. It is an open question whether it is also necessary (given Non-Collinearity).

We prove one direction here (sufficiency of (6.6)) and relegate the other direction to an appendix. For any belief function $\theta$, (6.6) implies that, for every $n$ and $f_n \in \mathcal{F}_{\leq n}$,

$$\int f_n d\overline{L}\left(\cdot\right) \geq \int V_{\theta^\infty}\left(f_n\right) d\mu.$$

Let $f$ and $f_n$ be as in the statement of the axiom. Then, using (6.3),

$$\begin{aligned}
U\left(f\right) &= \Sigma_{s_1^n}\overline{L}\left(s_1,...,s_n\right)U_n\left(f \mid s_1^n\right) \\
&= \Sigma_{s_1^n}\overline{L}\left(s_1,...,s_n\right)U_n\left(f_n \mid s_1^n\right) \\
&= \Sigma_{s_1^n}\overline{L}\left(s_1,...,s_n\right)f_n\left(s_1^n\right) \\
&\geq \left(\int V_{\theta^\infty}\left(f_n\right)d\mu\left(\theta\right)\right) = U\left(f_n\right),
\end{aligned}$$

where the inequality follows from the hypothesis (6.6).

Our final axiom is another expression of symmetry.

*Commutativity*: For all $n$ and samples $s_1^n$, acts $f', f \in \mathcal{F}_{>n}$, and finite permutations $\pi$,

$$f' \succeq_{n,s_1^n} f \implies f' \succeq_{n,\pi s_1^n} f,$$

where $\pi s_1^n$ is the permuted sample $\left(s_{\pi(1)},...,s_{\pi(n)}\right)$.

Since updating coincides with that in a Bayesian model with prior $\overline{L}$, Commutativity is satisfied if and only if conditionals of $\overline{L}$ are invariant to the order of the conditioning sample. A stronger, and hence sufficient condition, is that $\overline{L}$ be exchangeable.[25]

### 6.3. The Dynamics of Beliefs

Two properties are immediate: (i) Ambiguity is in general not monotonic along a sample. Posterior probabilities $\mu_n\left(\theta\right)$ are not monotonic under Bayesian updating. Thus, for example, if $\mu$ has two points of support $\theta'$ and $\theta$, and if $core(\theta') \supset core\left(\theta\right)$, then the set of priors corresponding to $U_n$ decreases with $n$ (in the sense of set inclusion) if $\mu_n\left(\theta\right)$ increases but increases in size if $\mu_n\left(\theta\right)$ decreases. (ii)

[25]See Fortini *et al* [17, pp. 90-2] for elaboration on how/why exchangeability is a strictly stronger condition.

Ambiguity need not vanish asymptotically (this is illustrated and discussed further below).[26]

To say more, and to illustrate what the model can deliver, we restrict the likelihood function in a way that seems natural but is only partly justified axiomatically. Assume that for ($\mu$-almost) every $\theta$:

- L1 $L(\cdot \mid \theta) \in core(\theta^\infty)$: This implies $\overline{L}(\cdot) \in core\left(\int \theta^\infty d\mu\right)$ and hence Payoff Ambiguity.
- L2 $L(\cdot \mid \theta)$ is exchangeable: This implies that $\overline{L}$ is exchangeable and hence Commutativity. By the de Finetti Theorem applied to $L(\cdot \mid \theta)$, the preceding is equivalent to $L(\cdot \mid \theta)$ being expressible in the form

$$L(\cdot \mid \theta) = \int_{\Delta(S)} \ell^\infty(\cdot)\, d\lambda_\theta(q), \tag{6.8}$$

  for some (unique) probability measure $\lambda_\theta$ on $\Delta(S)$.
- L3 $\lambda_\theta$ has support equal to $core(\theta)$: That the support is contained in $core(\theta)$ is implied by L1. Here we assume "full support."

The interpretation is as follows. There is ex ante uncertainty, represented by $\mu$, about which hypothesis $\theta$ best describes experiments. Conditional on a particular $\theta$ in the support of $\mu$, the individual is not confident in any particular probability law being accurate. For purposes of choice she is not Bayesian with respect to this uncertainty - ex ante ambiguity about each experiment is captured by the multiplicity of measures in $core(\theta^\infty)$, and her preferences on $\mathcal{F}$ are represented by the belief function utility $V_{\theta^\infty}$. When it comes to inference ex post, she is still uncertain about the data generating mechanism, but now she acts like a Bayesian - she uses the probability measure $\lambda_\theta$ over $core(\theta)$ to define likelihoods as in (6.8). The full support assumption L3 requires that she admit any measure in the core as a possibility for purposes of inference. If $\theta$ is additive, then L1 implies that $L(\cdot \mid p) = p^\infty(\cdot)$; therefore, if every $\theta$ in the support of $\mu$ is additive, 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.

[26] Though our model does not permit infinite samples, asymptotic results can be interpreted as an approximation to large finite samples.

**Proposition 6.4.** *Suppose that the likelihood function $L$ satisfies L1 and L2, and that $\mu$ has finite support. (i) Suppose further that for any $\theta' \neq \theta$ in the support, core $(\theta')$ and core $(\theta)$ are disjoint. Then, for every $\theta$ with $\mu(\theta) > 0$,*

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

*(ii) Let $\mu$ have support $\{\theta, p\}$, where $p \in$ core $(\theta)$ is permitted. If $p$ is not an atom of $\lambda_\theta$, that is, if $\lambda_\theta(\{p\}) = 0$, then*

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

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

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

Particularly if $\lambda_\theta$ has full support (L3), this clarifies the sense in which $G$ is a large set.

Note that even given certainty about $\theta$, in general there remains ambiguity when predicting future experiments and ranking bets over their outcomes. For example, in the coin-tossing example, an individual would become certain about the physical bias of the coin, but would (if $\theta$ is not additive) 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 joint probability measure $p^\infty$, if the truth has positive subjective probability ex ante ($\theta = p$ is in the support of $\mu$), 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 $p^\infty(\cdot)$ to predict future outcomes).[27]

Part (ii) is an illustrative result for the case when cores may overlap. Here there is convergence to the truth $p^\infty$, though the prior attaches positive probability 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 she sees each new coin-tosser as employing a different and hard-to-understand

[27] This is because, as indicated above, $L(\cdot \mid \theta) = p^\infty(\cdot)$ if $\theta = p$ is additive.

technique, then, even after learning the coin's bias, it is rational to take this limited understanding into account for further prediction and choice. In any case, the model does not force ambiguity to persist in all circumstances.

A final example exploits the fact that the "parameters" $\theta$ being learned about are belief functions, or, in the coin-tossing setting that we consider, each $\theta$ corresponds to a probability interval $I_\theta \subset [0,1]$. Let $\mu$ have support $\{\theta', \theta\}$, where

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

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 $\theta$. We ask how the posterior probability $\mu_n(\theta)$ behaves in large samples.

Specialize L1-L3 by assuming further that $\lambda_{\theta'}$ and $\lambda_\theta$ are uniform on their respective intervals. Though we do this for concreteness, the uniform distribution seems natural at a superficial level. It delivers the following result for the limiting probability of $\theta$:[28] Denote by $\Omega_\theta$ the set of samples $\omega$ for which $\lim \Psi_n(\omega) \in I_\theta$. Then, for every $\omega$ in $\Omega_\theta$,

$$\mu_\infty(\theta) = \frac{1}{1 + \frac{\mu(\theta')}{\mu(\theta)} \frac{\delta}{\delta'}}. \tag{6.9}$$

Note that, by (C.1) and the full support property L3, the set of samples $\Omega_\theta$ has positive probability according to both $L(\cdot \mid \theta')$ and $L(\cdot \mid \theta)$.

For samples in $\Omega_\theta$, the limiting empirical frequency of Heads is consistent with *both* $\theta$ and $\theta'$. This identification problem leads to the result that $0 < \mu_\infty(\theta) < 1$ - 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* [2]. They posit a prior of the form $\overline{L}(\cdot) = \int L(\cdot \mid \theta)\, d\mu(\theta)$ in order to model learning about a parameter when there is uncertainty about how to interpret a signal. Their model is Bayesian, and $\theta$ is an abstract parameter rather than a belief function. Here signals are difficult to interpret only because of the concern that experiments differ, while experiments are identical (in the behavioral sense of our paper) in their model.[29]

[28] The claim (6.9) to follow is adapted from [2, Lemma 1]. The latter implies also that for the lack of asymptotic learning, it would be enough for $\lambda_{\theta'}$ and $\lambda_\theta$ to have positive and continuous Lebesgue densities on their intervals.

[29] In spite of these differences, their results translate into our setting. In particular, one could use concern about nonidentical experiments to justify asymptotic disagreement between individuals.

Another noteworthy implication of (6.9) is that $\mu_{\infty}(\theta) > \mu(\theta)$, 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 $\theta'$ and $\theta$, but they differ more according to $\theta'$. Thus the sample provides less information about future experiments under $\theta'$ than under $\theta$. This leads to a shift in weight towards $\theta$.

# 7. RELATED LITERATURE

Shafer [39] is the first, to our knowledge, to discuss the use of belief functions within the framework of parametric statistical models analogous to de Finetti's. In particular, he sketches (section 3.3) a de Finetti-style treatment of randomness based on belief functions. His model is not axiomatic or choice-based, but ignoring these differences, one can translate his suggested model into our framework in the following way. Consider the de Finetti representation (1.1), where the probability measure $\mu$ models beliefs about $\ell$, the unknown 'parameter'. An obvious generalization is to replace $\mu$ by a belief function on $\Delta(S)$, or more generally by a set of probability measures on $\Delta(S)$, thus generalizing prior beliefs. Epstein and Seo [11] show that such a model is characterized primarily by an axiom called *Strong Exchangeability*, whereby randomizing between an act and any permuted variant of the act is of no value; formally,

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

The resulting representation is such that every prior in the set representing preference is exchangeable. This suggests the interpretation that there is prior ambiguity about the true law $\ell$, but, since the same $\ell$ applies to every experiment, there remains, just as in de Finetti's model, certainty that experiments are identical. For coin tossing, there is ambiguity about the given coin's bias, but the way in which the coin is tossed is viewed as fixed across experiments, and hence these are viewed as identical. It remains to formulate a general model that features *both* prior ambiguity *and* indistinguishable, but not necessarily identical, experiments. (See the concluding section of our other paper for a conjecture in this regard.)

Our paper [11] is evidently closely related. In addition to the representation just mentioned based on Strong Exchangeability, where experiments are identical, we also characterize a representation closer to the one in this paper, where experiments are not identical. This is done within the framework of multiple-priors utility. Since belief function utility is a subclass of the multiple-priors model, one might wonder whether the representation result Theorem 4.1 could be obtained simply by specializing the ambient preference framework and assuming that all preferences conform to belief function utility.

We emphasize that Theorem 4.1 is *not* a "special case," or simple corollary, of results in [11]. One reason is that the latter paper employs an axiom called Dominance that is redundant here.[30] There is an important difference also in the representations. The specialization of [11] indicated above would at best yield a representation wherein, though $\nu$ in (1.2) is assumed to be a belief function, the counterpart of each $\theta$ on the right-hand side would be the lower envelope of a set of priors, and not necessarily a belief function on $S$. The representation result obtained here is also sharper in another way. The rule for forming the i.i.d. product $\theta^{\infty}$ is pinned down - it corresponds to that advocated by Hendon *et al* [22]. In contrast, the representation in [11] is less specific in this regard because it reflects the well-known fact (see Ghirardato [19], in particular) that stochastic independence is more complicated in the multiple-priors (or nonadditive probability) framework in that there is more than one way to form independent products.

In addition, there is no counterpart in [11] of the connection established here (Corollary 5.1) between subjective prior beliefs and long run empirical frequencies, or of our model of updating.

In [11], we describe the connection to Epstein and Schneider [13] and especially [14]. The comparison with this paper suggests a trade-off between DC and Symmetry.[31] The (nonaxiomatic) model in [14] satisfies DC, but not Symmetry, while here we keep the latter at the cost of a weaker dynamic consistency property (WDC). The advantages of dynamic consistency for a normative model are evident. However, we find Symmetry compelling in a cross-sectional setting, and appealing also when experiments are separated in time. Therefore, in this paper, we have assumed Symmetry and explored a weaker, but still useful, form of dynamic consistency.

A recent paper by Al-Najjar and De Castro [1] also extends the de Finetti Theorem beyond subjective expected utility theory. Their key axiom builds on Strong Exchangeability and strengthens it to require that any act $f$ is indifferent

[30] Two more minor differences are: (i) Orthogonal Independence as stated here is considerably weaker than the corresponding axiom in our other paper; and (ii) Theorem 4.1 assumes that $S$ is compact metric, while in our other paper $S$ is required to be finite.

[31] See [12] for a formal result describing such a trade-off.

to any finite mixture of permutations of $f$, that is,

$$\Sigma_i \alpha_i (\pi_i f) \sim f.$$

In the multiple-priors framework, this axiom is equivalent to Strong Exchangeability, given also Symmetry. They offer results regarding the connection between preference and limiting empirical frequencies, with Theorem 1 being the main result. If one stays within the multiple-priors framework, then a counterpart of Theorem 1 is immediate once one has the representation in [11, Theorem 3.3], because as noted above, every prior in the set of priors is exchangeable and standard ergodic theory applies. Therefore, their analysis can be seen largely as a generalization of [11, Theorem 3.3] beyond the multiple-priors utility framework; for example, they can accommodate variational utility [28], and, since they do not impose completeness, also Bewley's [5] Knightian decision theory. (The latter, however, is readily done directly without any of the machinery in their paper or in ours.[32]) Finally, from the perspective of this paper (as opposed to that of [11]), and just as in Shafer's model, experiments are perceived to be identical in their model.

# A. Appendix: Proof of Theorem 4.1

First we prove the measurability required to show that the integrals in (4.2) and (4.3) are well-defined. (Recall that the Borel probability measure $\mu$ has a unique extension to the class of all universally measurable subsets.)

**Lemma A.1.** *Both $\theta \longmapsto V_{\theta^\infty}(f)$ and $\theta \longmapsto \theta^\infty(A)$ are universally measurable for any $f \in \mathcal{F}$ and $A \in \Sigma$.*

**Proof.** Since $Bel(S)$ and $\Delta(\mathcal{K}(S))$ are homeomorphic, and in light of (2.5), it is enough to prove analytical (and hence universal) measurability of the mapping from $\Delta(\mathcal{K}(S))$ to $\mathbb{R}$ given by

$$\ell \longmapsto \int_{[\mathcal{K}(S)]^\infty} \inf_{\omega \in K} f(\omega)\, d\ell^\infty(K).$$

[32]According to Bewley's model, $f \sim \pi f$ if and only if $Pf = (\pi P) f$ for every prior in the representing set of priors. Thus Symmetry implies that every prior in this set is exchangeable. This observation is due to Klaus Nehring.

Step 1. $\Delta(\mathcal{K}(S))$ and $\{\ell^\infty : \ell \in \Delta(\mathcal{K}(S))\}$ are homeomorphic when the latter set is endowed with the relative topology inherited from $\Delta([\mathcal{K}(S)]^\infty)$.

Step 2. $P \longmapsto \int \hat{f} dP$ from $\Delta([\mathcal{K}(S)]^\infty)$ to $\mathbb{R}$ is analytically measurable for any bounded analytically measurable function $\hat{f}$ on $[\mathcal{K}(S)]^\infty$: If $\hat{f}$ is simple (has a finite number of values), then $P \longmapsto \int \hat{f} dP$ is analytically measurable by [4, p. 169]. More generally, $\int \hat{f} dP$ equals the pointwise limit of $\lim \int \hat{f}_n dP$ for some simple and analytically measurable $\hat{f}_n$, which implies the desired measurability.

Step 3. Note that

$$\left\{K \in \mathcal{K} : \inf_{\omega \in K} f(\omega) \geq t\right\} = \{K \in \mathcal{K} : K \subset \{\omega : f(\omega) \geq t\}\} \tag{A.1}$$

is co-analytic by [35, p. 772], and hence analytically measurable.

Steps 1, 2 and 3 complete the proof. ■

For Theorem 4.1, we show (iii)⇒(ii)⇒(i)⇒(iii). If $\nu \in Bel$, let $m = \zeta(\nu)$. We use (2.5) repeatedly without reference.

(iii)⇒(ii): Let $\Sigma'$ be the $\sigma$-algebra generated by the class

$$\{K \in \mathcal{K} : K \subset A\}_{A \in \Sigma}.$$

We claim that $m(\cdot) = \int_{Bel(S)} \zeta(\theta^\infty)(\cdot)\, d\mu(\theta)$ on $\Sigma'$. Since the latter is a probability measure on $\mathcal{K}$, it is enough to show that

$$m(\{K \in \mathcal{K} : K \subset A\}) = \int_{Bel(S)} \zeta(\theta^\infty)(\{K \in \mathcal{K} : K \subset A\})\, d\mu(\theta)$$

for each $A \in \Sigma$. This is equivalent to

$$\nu(A) = \int_{Bel(S)} \theta^\infty(A)\, d\mu(\theta),$$

which is true given (iii).

By a standard argument using the Lebesgue Dominated Convergence Theorem,

$$\int_{\mathcal{K}} \hat{f} dm = \int_{Bel(S)} \left(\int_{\mathcal{K}} \hat{f} d\zeta(\theta^\infty)\right) d\mu(\theta),$$

for all $\Sigma'$-measurable $\hat{f} : \mathcal{K} \to [0,1]$. Since $K \longmapsto \inf_{\omega \in K} f(\omega)$ is $\Sigma'$-measurable by (A.1),

$$\begin{aligned} U_{\nu}(f) &= \int_{\mathcal{K}} \inf_{\omega \in K} f(\omega)\, dm(K) = \int_{Bel(S)} \left( \int_{\mathcal{K}} \inf_{\omega \in K} f(\omega)\, d\zeta(\theta^{\infty}) \right) d\mu(\theta) \\ &= \int_{Bel(S)} V_{\theta^{\infty}}(f)\, d\mu(\theta). \end{aligned}$$

(ii)⇒(i): It is enough to show that $V_{\theta^{\infty}}$ satisfies Symmetry and OI. Let $m = \zeta(\theta^{\infty}) = (\zeta(\theta))^{\infty}$. Then, $m$ is an i.i.d. measure on $[\mathcal{K}(S)]^{\infty}$. Since $m$ is symmetric,

$$\begin{aligned} V_{\theta^{\infty}}(\pi f) &= \int_{\mathcal{K}} \inf_{\omega \in K} \pi f(\omega)\, dm(K) = \int_{\mathcal{K}} \inf_{\omega \in K} f(\pi\omega)\, dm(K) \\ &= \int_{\mathcal{K}} \inf_{\pi\omega \in \pi K} f(\pi\omega)\, dm(K) = \int_{\mathcal{K}} \inf_{\omega \in K} f(\omega)\, d(\pi m)(K) \\ &= \int_{\mathcal{K}} \inf_{\omega \in K} f(\omega)\, dm(K) = V_{\theta^{\infty}}(f). \end{aligned}$$

Show (3.2) to prove OI. For simplicity, let $f \in \mathcal{F}_1$ and $g \in \mathcal{F}_2$. The general case is similar. For $0 < \alpha \leq 1$,

$$\begin{aligned} & V_{\theta^{\infty}}(\alpha f + (1-\alpha) g) \\ = & \int_{\mathcal{K}} \inf_{\omega \in K} [\alpha f(\omega) + (1-\alpha) g(\omega)]\, dm(K) \\ = & \int_{[\mathcal{K}(S)]^{\infty}} \inf_{s_1 \in K_1, s_2 \in K_2} [\alpha f(s_1) + (1-\alpha) g(s_2)]\, dm(K_1, K_2, ...) \end{aligned}$$

$$\begin{aligned} = & \int_{[\mathcal{K}(S)]^{\infty}} \alpha \left[ \inf_{s_1 \in K_1} f(s_1) \right] + (1-\alpha) \left[ \inf_{s_2 \in K_2} (1-\alpha) g(s_2) \right] dm(K_1, K_2, ...) \\ = & \alpha \int_{[\mathcal{K}(S)]^{\infty}} \left[ \inf_{s_1 \in K_1} f(s_1) \right] dm(K_1, K_2, ...) \\ & + (1-\alpha) \int_{[\mathcal{K}(S)]^{\infty}} \left[ \inf_{s_2 \in K_2} g(s_2) \right] dm(K_1, K_2, ...) \\ = & \alpha V_{\theta^{\infty}}(f) + (1-\alpha) V_{\theta^{\infty}}(g). \end{aligned}$$

The second equality follows because $K \in [\mathcal{K}(S)]^{\infty}$, $a.s.$-$m[K]$.

(i)$\Rightarrow$(iii): For $C \subset \mathcal{K}$, let $\pi C = \{\pi K \in \mathcal{K} : K \in C\}$, and for $m \in \Delta(\mathcal{K})$, define $\pi m \in \Delta(\mathcal{K})$ by $\pi m(C) = m(\pi C)$ for each $C \in \Sigma_{\mathcal{K}}$.

**Lemma A.2.** *For any $m \in \Delta(\mathcal{K})$, $m = \pi m$ for all $\pi$ if and only if $m = \zeta(\nu)$ for some symmetric belief function $\nu$ on $\Omega$.*

**Proof.** If $m = \zeta(\nu)$, then $\nu(K) = m(\{K' \in \mathcal{K} : K' \subset K\})$, and

$$\begin{aligned} \nu(\pi K) &= m(\{K' \in \mathcal{K} : K' \subset \pi K\}) = m(\{\pi K' \in \mathcal{K} : \pi K' \subset \pi K\}) \\ &= m(\{\pi K' \in \mathcal{K} : K' \subset K\}) = m(\pi(\{K' \in \mathcal{K} : K' \subset K\})). \end{aligned}$$

The asserted equivalence follows, because the class $\{K' \in \mathcal{K} : K' \subset K\}_{K \in \mathcal{K}}$ generates the Borel $\sigma$-algebra on $\mathcal{K}$. ■

**Lemma A.3.** *Let $\nu$ be a belief function on $S^{\infty}$ and $m = \zeta(\nu)$ the corresponding measure on $\mathcal{K}(S^{\infty})$. If $U_{\nu}$ satisfies OI, then $m[(\mathcal{K}(S))^{\infty}] = 1$.*

**Proof.** For any $\omega \in S^{\infty}$ and disjoint sets $I, J \subset \mathbb{N}$, $\omega_I$ denotes the projection of $\omega$ onto $S^I$, and we write $\omega = (\omega_I, \omega_J, \omega_{-I-J})$. When $I = \{i\}$, we write $\omega_i$, rather than $\omega_{\{i\}}$, to denote the $i$-th component of $\omega$.

Let $\mathcal{A}$ be the collection of compact subsets $K$ of $S^{\infty}$ satisfying: For any $n > 0$, and $\omega^1, \omega^2 \in K$, and for every partition $\{1, ..., n\} = I \cup J$,

$$\exists \omega^* \in K, \text{ such that } \omega^*_I = \omega^1_I \text{ and } \omega^*_J = \omega^2_J. \tag{A.2}$$

In other words, for every $n$, the projection of $K$ onto $S^n$ is a Cartesian product.

*Step 1.* For any *continuous* acts $f \in \mathcal{F}_I$ and $g \in \mathcal{F}_J$ with finite disjoint $I$ and $J$,

$$\min_{\omega \in K} \left[\tfrac{1}{2} f(\omega) + \tfrac{1}{2} g(\omega)\right] = \tfrac{1}{2} \min_{\omega \in K} f(\omega) + \tfrac{1}{2} \min_{\omega \in K} g(\omega), \tag{A.3}$$

$a.s.$-$m[K]$: This is where OI enters - by (3.2) it implies that

$$U_{\nu}\left(\tfrac{1}{2} f + \tfrac{1}{2} g\right) = \tfrac{1}{2} U_{\nu}(f) + \tfrac{1}{2} U_{\nu}(g).$$

Since $U_\nu(f) = \int_{\mathcal{K}} \inf_{\omega\in K} f(\omega)\, dm(K)$,

$$\int_{\mathcal{K}} \min_{\omega\in K}\left[\tfrac{1}{2}f(\omega)+\tfrac{1}{2}g(\omega)\right] dm(K) = \tfrac{1}{2}\int_{\mathcal{K}} \min_{\omega\in K} f(\omega)\, dm(K) + \tfrac{1}{2}\int_{\mathcal{K}} \min_{\omega\in K} g(\omega)\, dm(K).$$

The assertion follows from

$$\min_{\omega\in K}\left[\tfrac{1}{2}f(\omega)+\tfrac{1}{2}g(\omega)\right] \geq \tfrac{1}{2}\min_{\omega\in K} f(\omega) + \tfrac{1}{2}\min_{\omega\in K} g(\omega).$$

Let $\mathcal{G}$ be the set of all pairs $(f,g)$ such that $f$ and $g$ are continuous and $f \in \mathcal{F}_I$, $g \in \mathcal{F}_J$ for some finite disjoint $I$ and $J$. Let $\mathcal{B}_{f,g}$ be the collection of $K \in \mathcal{K}$ satisfying (A.3), given $f$ and $g$. Step 1 implies $m(\mathcal{B}_{f,g}) = 1$ for each $(f,g) \in \mathcal{G}$.

*Step 2.* $m\left(\bigcap_{(f,g)\in\mathcal{G}} \mathcal{B}_{f,g}\right) = 1$: Since the set of continuous finitely-based acts is separable under the sup-norm topology (see [3, Lemma 3.99]), it is easy to see that $\mathcal{G}$ is also separable. Let $\{(f_n, g_n)\}$ be a countable dense subset of $\mathcal{G}$. By Step 1,

$$m\left(\mathcal{K}\backslash\left(\bigcap_{i=1}^{\infty}\mathcal{B}_{f_i,g_i}\right)\right) = m\left(\bigcup_{i=1}^{\infty}\left(\mathcal{K}\backslash\mathcal{B}_{f_i,g_i}\right)\right) \leq \sum m\left(\mathcal{K}\backslash\mathcal{B}_{f_i,g_i}\right) = 0.$$

Thus it is enough to show that $\bigcap_{i=1}^{\infty}\mathcal{B}_{f_i,g_i} = \bigcap_{(f,g)\in\mathcal{G}}\mathcal{B}_{f,g}$.

Only $\subset$ requires proof. Let $K \in \bigcap_{i=1}^{\infty}\mathcal{B}_{f_i,g_i}$, $(f,g) \in \mathcal{G}$ and assume wlog that $(f_i, g_i) \to (f,g)$. Then, by the Maximum Theorem [3, Theorem 17.31],

$$\begin{aligned}
\min_{\omega\in K}\left[\tfrac{1}{2}f(\omega)+\tfrac{1}{2}g(\omega)\right] &= \lim_i \min_{\omega\in K}\left[\tfrac{1}{2}f_i(\omega)+\tfrac{1}{2}g_i(\omega)\right] \\
&= \lim_i \left[\tfrac{1}{2}\min_{\omega\in K} f_i(\omega) + \tfrac{1}{2}\min_{\omega\in K} g_i(\omega)\right] \\
&= \tfrac{1}{2}\min_{\omega\in K} f(\omega) + \tfrac{1}{2}\min_{\omega\in K} g(\omega).
\end{aligned}$$

Thus $K \in \bigcap_{(f,g)\in\mathcal{G}}\mathcal{B}_{f,g}$.

*Step 3.* If $K \in \bigcap_{(f,g)\in\mathcal{G}} \mathcal{B}_{f,g}$, then $K \in \mathcal{A}$: Let $n \geq 0$, $\omega^1, \omega^2 \in K$ and $\{1, ..., n\} = I \cup J$, with $I$ and $J$ disjoint. For each $i$, take closed sets

$$\begin{aligned} A_i &= \left\{ \omega : \sum_{t\in I} 2^{-t} d\left(\omega_t, \omega_t^1\right) \geq \frac{1}{i} \right\} \text{ and} \\ B_i &= \left\{ \omega : \sum_{t\in J} 2^{-t} d\left(\omega_t, \omega_t^2\right) \geq \frac{1}{i} \right\}, \end{aligned}$$

where $d(\cdot, \cdot)$ is the metric on $S$. By Urysohn's Lemma, there are continuous functions $f_i$ and $g_i$ such that, for each $i$,

$$\begin{aligned} f_i(\omega) &= 1 \text{ if } \omega \in A_i \text{ and } 0 \text{ if } \omega_I = \omega_I^1, \text{ and} \\ g_i(\omega) &= 1 \text{ if } \omega \in B_i \text{ and } 0 \text{ if } \omega_J = \omega_J^2. \end{aligned}$$

Since $A_i \in \Sigma_I$ and $B_i \in \Sigma_J$, we can take $f_i \in \mathcal{F}_I$, and $g_i \in \mathcal{F}_J$. Then, $\min_{\omega\in K} f_i(\omega) = \min_{\omega\in K} g_i(\omega) = 0$ and, since $K \in \mathcal{B}_{f_i,g_i}$,

$$\min_{\omega\in K} [f_i(\omega) + g_i(\omega)] = 0.$$

Hence, there exists $\hat{\omega}^i \in K$ such that $f_i\left(\hat{\omega}^i\right) = g_i\left(\hat{\omega}^i\right) = 0$. By the construction of $f_i$ and $g_i$, we have $\hat{\omega}^i \notin A_i, B_i$, which implies

$$\sum_{t\in I} 2^{-t} d\left(\hat{\omega}_t^i, \omega_t^1\right) + \sum_{t\in J} 2^{-t} d\left(\hat{\omega}_t^i, \omega_t^2\right) < \frac{2}{i}.$$

Since $\{\hat{\omega}^i\} \subset K$ and $K$ is compact, there is a limit point $\omega^* \in K$ satisfying (A.2).

*Step 4.* $m(\mathcal{A}) = 1$: By Steps 2-3, $1 \geq m(\mathcal{A}) \geq m\left(\bigcap_{(f,g)\in\mathcal{G}} \mathcal{B}_{f,g}\right) = 1.$

*Step 5.* $\mathcal{A} = (\mathcal{K}(S))^\infty$: Clearly $\mathcal{A} \supset (\mathcal{K}(S))^\infty$. For the other direction, take $K \in \mathcal{A}$ and assume $\omega^1, \omega^2, ... \in K$. It suffices to show that

$$\omega^* = \left(\omega_1^1, \omega_2^2, ..., \omega_n^n, ...\right) \in K. \tag{A.4}$$

Since $K \in \mathcal{A}$ and $\omega^1, \omega^2 \in K$, there exists $\hat{\omega}^2 \in K$ such that $\left(\hat{\omega}_1^2, \hat{\omega}_2^2\right) = (\omega_1^1, \omega_2^2)$. Similarly, since $\hat{\omega}^2, \omega^3 \in K$, there exists $\hat{\omega}^3 \in K$ such that $\left(\hat{\omega}_1^3, \hat{\omega}_2^3, \hat{\omega}_3^3\right) =$

$\left(\hat{\omega}_1^2, \hat{\omega}_2^2, \omega_3^3\right) = (\omega_1^1, \omega_2^2, \omega_3^3)$, and so on, giving a sequence $\{\hat{\omega}^n\}$ in $K$. Any limit point $\omega^*$ satisfies (A.4). ■

Finally, we prove (i)⇒(ii). Let $\nu$ be a belief function on $S^\infty$ and suppose that $U_\nu$ satisfies Symmetry and OI. By Lemma A.3, $m \equiv \zeta(\nu)$ can be viewed as a measure on $[\mathcal{K}(S)]^\infty$, and by Lemma A.2, $m$ is symmetric. Thus we can apply de Finetti's Theorem [23] to $m$, viewing $\mathcal{K}(S)$ as the one-period state space, to obtain: There exists $\hat{\mu} \in \Delta(\Delta(\mathcal{K}(S)))$ such that

$$m(C) = \int_{\Delta(\mathcal{K}(S))} \ell^\infty(C)\, d\hat{\mu}(\ell) \text{ for all } C \in \Sigma_{[\mathcal{K}(S)]^\infty}.$$

Here each $\ell$ lies in $\Delta(\mathcal{K}(S))$ and $\ell^\infty$ is the i.i.d. product measure on $[\mathcal{K}(S)]^\infty$. Extend each measure $\ell^\infty$ to $\Sigma_{\mathcal{K}}$ and write

$$m(C) = \int_{\Delta(\mathcal{K}(S))} \ell^\infty(C)\, d\hat{\mu}(\ell) \text{ for all } C \in \Sigma_{\mathcal{K}(S^\infty)}.$$

We claim that the equation extends also to $C \in \Sigma'$, where $\Sigma'$ is the $\sigma$-algebra generated by the class

$$\{K \in \mathcal{K} : K \subset A\}_{A \in \Sigma}.$$

First, note that $\ell \longmapsto \ell^\infty(C)$ is universally measurable by Lemma A.1, and hence the integral is well-defined. By a standard argument using the Lebesgue Dominated Convergence Theorem, $C \longmapsto \int_{\Delta(\mathcal{K}(S))} \ell^\infty(C)\, d\hat{\mu}(\ell)$ is countably additive on $\Sigma'$. This completes the argument because $m$ has a unique extension to the $\sigma$-algebra of universally measurable sets, and the latter contains $\Sigma'$.

Let $\mu \equiv \hat{\mu} \circ \zeta \in \Delta(Bel(S))$ and apply the Change of Variables Theorem to derive, for any $A \in \Sigma$,

$$\begin{aligned} \nu(A) &= m(\{K \in \mathcal{K} : K \subset A\}) \\ &= \int_{\Delta(\mathcal{K}(S))} \ell^\infty(\{K \in \mathcal{K} : K \subset A\})\, d\hat{\mu}(\ell) \\ &= \int_{\Delta(\mathcal{K}(S))} \ell^\infty(\{K \in \mathcal{K} : K \subset A\})\, d\mu \circ \zeta^{-1}(\ell) \end{aligned}$$

$$
\begin{aligned}
&= \int_{Bel(S)} [\zeta(\theta)]^{\infty} (\{K \in \mathcal{K} : K \subset A\})\, d\mu(\theta) \\
&= \int_{Bel(S)} \zeta(\theta^{\infty}) (\{K \in \mathcal{K} : K \subset A\})\, d\mu(\theta) \\
&= \int_{Bel(S)} \theta^{\infty}(A)\, d\mu(\theta).
\end{aligned}
$$

Uniqueness of $\mu$ follows from the uniqueness of $\hat{\mu}$ provided by de Finetti's Theorem. ■

# B. Appendix: Proofs for Section 5

*Proof of Corollary 5.1*: Since $\theta^{\infty}(A) = \theta(A)$ for $A \in \Sigma_S$, $\theta \longmapsto \theta(A)$ is universally measurable by Lemma A.1. Hence, every set of the form

$$
\{\theta \in Bel(S) : [\theta(A), 1 - \theta(S \backslash A)] \subset [a, b]\}
$$

is universally measurable and the statement of the Corollary is well-defined.

We need two lemmas. Recall that $\Psi_n(A)(\omega) = \frac{1}{n}\sum_{i=1}^{n} I(\omega_i \in A)$ where $\omega_i$ is the $i$-th component of $\omega$. Similarly define $\widehat{\Psi}_n(A)(K) = \frac{1}{n}\sum_{i=1}^{n} I(K_i \subset A)$ for $K \in [\mathcal{K}(S)]^{\infty}$, where $K_i$ is the $i$-th component of $K$.

**Lemma B.1.** *Let $K \in [\mathcal{K}(S)]^{\infty}$, $K = K_1 \times K_2 \times \ldots$ , and $\alpha \in \mathbb{R}$. Then the following are equivalent:*
*(i)* $\liminf_n \Psi_n(A)(\omega) > \alpha$ *for every* $\omega_i \in K_i$, $i = 1, \ldots$
*(ii)* $\liminf_n \widehat{\Psi}_n(A)(K) > \alpha$.

**Proof.** (i)⇒(ii): If $K_i \subset A$, let $\omega_i$ be any element in $K_i$, and otherwise, let $\omega_i$ be any element in $K_i \backslash A$. Then, $I(K_i \subset A) = I(\omega_i \in A)$ and thus (ii) is implied.

(ii)⇒(i): If $\omega_i \in K_i$, $I(K_i \subset A) \leq I(\omega_i \in A)$. Thus, if $\omega_i \in K_i$ for $i = 1, \ldots$, then,

$$
\liminf_n \Psi_n(A)(\omega) \geq \liminf_n \widehat{\Psi}_n(A)(K) > \alpha.
$$

■

**Lemma B.2.** *(i)* $\theta^{\infty}(\{\omega : \theta(A) < \liminf_n \Psi_n(A)(\omega)\}) = 0$ *for each* $A \in \Sigma_S$; *and*
*(ii)* $\theta^{\infty}(\{\omega : \limsup_n \Psi_n(A)(\omega) < 1 - \theta(S \backslash A)\}) = 0$ *for each* $A \in \Sigma_S$.

**Proof.** Fix $A \in \Sigma_S$. Then,

$$\begin{aligned}
& \theta^{\infty}\left(\left\{\omega : \theta(A) < \liminf_n \Psi_n(A)(\omega)\right\}\right) \\
= \; & [\zeta(\theta)]^{\infty}\left(\left\{K \in [\mathcal{K}(S)]^{\infty} : K \subset \left\{\omega : \theta(A) < \liminf_n \Psi_n(A)(\omega)\right\}\right\}\right) \\
= \; & [\zeta(\theta)]^{\infty}\left(\left\{K \in [\mathcal{K}(S)]^{\infty} : \liminf_n \widehat{\Psi}_n(A)(K) > \theta(A)\right\}\right) \quad \text{(by Lemma B.1).}
\end{aligned}$$

By the Law of Large Numbers, $\widehat{\Psi}_n(A)(K)$ converges to $\zeta(\theta)(\{K_1 \in \mathcal{K}(S) : K_1 \subset A\}) = \theta(A)$ almost surely-$[\zeta(\theta)]^{\infty}$, which implies (i). The proof of (ii) is similar. ■

Return to the Corollary. By the LLN in [27], Lemma B.2 and the monotonicity of belief functions,

$$\begin{aligned}
& \theta^{\infty}(\{\omega : [\liminf \Psi_n(A)(\omega), \limsup \Psi_n(A)(\omega)] \subset [a,b]\}) = 1 \\
\Leftrightarrow \; & [\theta(A), 1 - \theta(S \backslash A)] \subset [a,b]
\end{aligned}$$

and

$$\begin{aligned}
& \theta^{\infty}(\{\omega : [\liminf \Psi_n(A)(\omega), \limsup \Psi_n(A)(\omega)] \subset [a,b]\}) = 0 \\
\Leftrightarrow \; & [\theta(A), 1 - \theta(S \backslash A)] \text{ is not a subset of } [a,b].
\end{aligned}$$

Moreover, for any belief function $\gamma$ on $\Omega$, if $\gamma(A) = \gamma(B) = 1$, then $\gamma(A \cap B) = 1$ by the Choquet theorem. Therefore,

$$\begin{aligned}
& \nu\left(\bigcap_{i=1}^{I} \{\omega : [\liminf \Psi_n(A_i)(\omega), \limsup \Psi_n(A_i)(\omega)] \subset [a_i, b_i]\}\right) \\
= \; & \int_{Bel(S)} \theta^{\infty}\left(\bigcap_{i=1}^{I} \{\omega : [\liminf \Psi_n(A_i)(\omega), \limsup \Psi_n(A_i)(\omega)] \subset [a_i, b_i]\}\right) d\mu(\theta) \\
= \; & \mu\left(\bigcap_{i=1}^{I} \{\theta : [\theta(A_i), 1 - \theta(S \backslash A_i)] \subset [a_i, b_i]\}\right). \qquad ■
\end{aligned}$$

*Proof of Proposition 5.2*: By exploiting the homeomorphism defined in the Choquet Theorem, we can identify $\mu'$ and $\mu$ with measures on $\Delta(\mathcal{K}(S))$. Modulo this identification, we are given that $\mu'$ and $\mu$ agree on the collection of all subsets of $\Delta(\mathcal{K}(S))$ of the form

$$\bigcap_{i=1}^{I} \{\ell \in \Delta(\mathcal{K}(S)) : \ell(\{K \in \mathcal{K}(S) : K \subset A_i\}) \geq a_i\},$$

for all $I > 0$, $A_i \in \Sigma_S$ and $a_i \in [0,1]$. They necessarily agree also on the generated $\sigma$-algebra, denoted $\Sigma^*$. Therefore, it suffices to show that

$$\Sigma_{\Delta(\mathcal{K}(S))} \subset \Sigma^*,$$

where $\Sigma_{\Delta(\mathcal{K}(S))}$ is the Borel $\sigma$-algebra on $\Delta(\mathcal{K}(S))$.

Step 1. $\ell \longmapsto \ell(C)$ is $\Sigma^*$-measurable for measurable $C \in \Sigma_{\mathcal{K}(S)}$: Let $\mathcal{C}$ be the collection of measurable subsets $C$ of $\mathcal{K}(S)$ such that $\ell \longmapsto \ell(C)$ is $\Sigma^*$-measurable. Every set of the form $\{K' \in \mathcal{K}(S) : K' \subset K\}$ for $K \in \mathcal{K}(S)$ lies in $\mathcal{C}$. Since the collection $\{K' \in \mathcal{K}(S) : K' \subset K\}_{K \in \mathcal{K}(S)}$ generates $\Sigma_{\mathcal{K}(S)}$, it is enough to show that $\mathcal{C}$ is a $\sigma$-algebra: (i) $C \in \mathcal{C}$ implies $\mathcal{K}(S) \backslash C \in \mathcal{C}$; (ii) if each $C_i \in \mathcal{C}$, then $\ell \longmapsto \ell(\cup_{i=1}^{\infty} C_i)$ is $\Sigma^*$-measurable because it equals the pointwise limit of $\ell \longmapsto \ell(\cup_{i=1}^{n} C_i)$ - hence $\cup_{i=1}^{\infty} C_i \in \mathcal{C}$.

Step 2. $\ell \longmapsto \int \hat{f} d\ell$ is $\Sigma^*$-measurable for all Borel-measurable $\hat{f}$ on $\mathcal{K}(S)$: Identical to Step 2 in Lemma A.1.

Step 3. $\Sigma_{\Delta(\mathcal{K}(S))} \subset \Sigma^*$: By Step 2, $\left\{\ell : \int \hat{f} d\ell \geq a\right\} \in \Sigma^*$ for all Borel-measurable $\hat{f}$ on $\mathcal{K}(S)$. But $\Sigma_{\Delta(\mathcal{K}(S))}$ is the smallest $\sigma$-algebra containing the sets $\left\{\ell : \int \hat{f} d\theta \geq a\right\}$ for all continuous $\hat{f}$ and $a \in \mathbb{R}$. ■

# C. 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 [29], for which the main step is to show that $D$ is convex, where

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

A preliminary result concerns shifted acts. Denote by $\varkappa$ the *shift* operator, so that, for any act,

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

$\varkappa^n$ denotes the $n$-fold replication of $\varkappa$. We show in [11, Lemma 3.8] that Symmetry implies also indifference to shifts, that is, $\varkappa f \sim f$ for all acts $f$.

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 $\varkappa^N g$ are orthogonal, that is, they depend on disjoint sets of experiments. For such an $N$, because each utility function satisfies OI and shift-invariance,

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

where the last equality follows from (3.2). Finally, the preceding can be extended to general (not only finitely-based) acts $f$ and $g$. That is because, as in our other model [11], preference is "regular" and thus properties on the set of finitely-based acts extend to all acts. We refer the reader to our earlier paper for elaboration and more precise statements. That belief function utility is regular follows from [35, Proposition 1] and [11, Theorem 2.2].

The other conditions in Proposition 1 of [29] are readily verified.[33] 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 $f \in \mathcal{F}_{>1}$,

$$\int V_{\theta^\infty}(f)\, d\mu(\theta) = \Sigma_{s_1} a_{s_1} \int V_{\theta^\infty}(f)\, d\mu_{s_1}(\theta) = \int V_{\theta^\infty}(f) \left(\Sigma_{s_1} a_{s_1} d\mu_{s_1}(\theta)\right).$$

By uniqueness of the representing measure,

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

[33] 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 $0 \leq p \leq 1$. 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

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

Similarly for $n > 1$.

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

**Remark 1.** *The above proof uses finiteness of $S$. An extension to an infinite state space may be possible, for example, by adapting Zhou's [42] proof of the Harsanyi theorem for infinite societies.*

*Proof of Corollary 6.3*: It remains to prove necessity of (6.6): Fix $f \in \mathcal{F}_{>n}$ and define $f_n \in \mathcal{F}_{\leq n}$ by

$$f_n(s_1^n) = U_n(f \mid s_1^n), \text{ for every } s_1^n.$$

By Payoff Ambiguity, $f \succeq f_n$, and hence, using (6.3),

$$\begin{aligned} \int f_n d\overline{L}(\cdot) &= U(f) \geq U(f_n) = \\ \int V_{\theta^\infty}(f_n)\, d\mu &= \int V_{\theta^n}(f_n)\, d\mu, \end{aligned}$$

or

$$\int f_n d\overline{L}(\cdot) \geq V_{\int \theta^n d\mu}(f_n).$$

Moreover, this inequality is valid for every $f_n \in D_n = \cup_{f \in \mathcal{F}_{>n}} \{f_n \in \mathcal{F}_{\leq n} : f_n(\cdot) = U_n(f \mid \cdot)\}$. Under Non-Collinearity, $D_n = \mathcal{F}_{\leq n}$: Identify $D_n$ with a subset of $[0,1]^{S^n}$. It is convex - see the proof of Theorem 6.1 - and includes the main diagonal (the portion within the unit cube). Non-Collinearity implies that $D_n$ is not contained within any hyperplane. Therefore, $D_n = [0,1]^{S^n}$, and,

$$mrg_{S_1 \times \ldots \times S_n} \overline{L}(\cdot) \in core\left( \int \theta^n d\mu \right).$$

It follows by standard arguments, using the regularity property Bel.4 of belief functions and the Choquet's [8, 15.2] capacitability result, that

$$\overline{L}(\cdot) \in core\left(\int \theta^{\infty} d\mu\right). \quad \blacksquare$$

*Proof of Proposition 6.4*: (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. To prove the general case, simply replace probability intervals by cores.

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

$$\lambda_\theta(I) = L(\cdot \mid \theta)(\{\omega : \lim \Psi_n(\omega) \in I\}). \tag{C.1}$$

Since $\lambda_\theta$ has support in $I_\theta \equiv [\theta(H), 1 - \theta(T)]$,

$$\lambda_\theta(I_\theta) = 1.$$

Because intervals are disjoint, for each $\omega$, there is at most one $\theta$ such that $\lim \Psi_n(\omega) \in I_\theta$. Define $F : \Omega \to Supp(\mu)$, by

$$F(\omega) = \theta, \text{ if } \lim \Psi_n(\omega) \in I_\theta,$$

and define $F(\omega) = \overline{\theta}$, with $\overline{\theta}$ an arbitrary fixed belief function in the support of $\mu$, if $\lim \Psi_n(\omega) \notin \cup_{Supp(\mu)} I_\theta$. Then,

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

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

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

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

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

The former is valid because $L(\{\omega : \lim \Psi_n(\omega) = p\} \mid \theta) = \lambda_\theta(\{p\}) = 0$. Thus [26, Proposition 3] completes the proof. $\blacksquare$

## References


[1] N. Al-Najjar and L.De Castro, A subjective foundation of objective probability, March 2009.

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

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

[4] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[5] T. Bewley, Knightian decision theory: Part I, *Decisions in Economics and Finance* 25 (2002), 79-110.

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

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

[8] G. Choquet, Theory of capacities, *Ann. Inst. Fourier* 5 (1954), 131-295.

[9] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Statist.* 38 (1967), 325-339.

[10] L.G. Epstein, M. Marinacci and K. Seo, Coarse contingencies and ambiguity, *Theor. Econ.* 2 (2007), 355-394.

[11] L.G. Epstein and K. Seo, Symmetry of evidence without evidence of symmetry, 2008.

[12] L.G. Epstein and K. Seo, Symmetry or dynamic consistency?, 2009.

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

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

[15] L.G. Epstein and T. Wang, Intertemporal asset pricing under Knightian uncertainty, *Econometrica* 62 (1994), 283-322.

[16] 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.

[17] S. Fortini, L. Ladelli and E. Regazzini, Exchangeability, predictive distributions and parametric models, *Indian J. Stats.* 62 (2000), 86-109.

[18] P. Ghirardato, Coping with ignorance, unforeseen contingencies and nonadditive uncertainty, *Econ. Theory* 17 (2001), 247-276.

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

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

[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] P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[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] F. Maccheroni and M. Marinacci, A strong law of large numbers for capacities, *Ann. Prob.* 33 (2005), 1171-1178.

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

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

[30] I. Molchanov, *Theory of Random Sets*, Springer, 2005.

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

[32] S. Mukerji, Understanding the nonadditive probability decision model, *Econ. Theory* 9 (1997), 23-46.

[33] H.T. Nguyen, On random sets and belief functions, *J. Math. Anal. and Appl.* 65 ((1978), 531-542.

[34] R. Phelps, *Lectures on Choquet's Theorem*, Springer, 2001.

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

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

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

[38] G. Shafer, *A Mathematical Theory of Evidence*, Princeton U. Press, Princeton, 1976.

[39] G. Shafer, Belief functions and parametric models (with commentary), *J. Royal Stat. Soc. Series B* 44 (1982), 322-352.

[40] M. Talagrand, Capacités invariantes extrémales, *Ann. Inst. Fourier* (Grenoble) 28 (1978), 79-146.

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

[42] L. Zhou, Harsanyi's utilitarianism theorems: general societies, *J. Econ. Theory* 72 (1997), 198-207.