



Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 157 (2015) 879–917

JOURNAL OF
**Economic
Theory**

www.elsevier.com/locate/jet

Exchangeable capacities, parameters and incomplete theories [☆]

Larry G. Epstein ^a, Kyoungwon Seo ^{b,*}

^a *Boston University, United States*

^b *Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea*

Received 25 January 2014; final version received 12 February 2015; accepted 20 February 2015

Available online 24 February 2015

Abstract

The de Finetti Theorem on exchangeable predictive priors is generalized to a framework where preference is represented by Choquet expected utility with respect to a belief function (a special capacity). The resulting model provides behavioral foundations for the decision-maker's subjective theory of the environment in which there are factors common to all experiments (or sources of uncertainty), called parameters, but in which her theory is incomplete in that knowledge of the parameter leaves idiosyncratic factors that vary across experiments in a way that is poorly understood.

© 2015 Elsevier Inc. All rights reserved.

JEL classification: D81

Keywords: Ambiguity; Exchangeability; Maxmin expected utility; Parameters; Repeated experiments

[☆] We gratefully acknowledge the financial support of the National Science Foundation, awards SES-0917740, SES-0918248 and SES-1216339. Some of the material in this paper was formerly contained in a paper titled “Bayesian inference and non-Bayesian prediction and choice ...”.

* Corresponding author.

E-mail addresses: lepstein@bu.edu (L.G. Epstein), kseo@kaist.ac.kr (K. Seo).

1. Introduction

1.1. Outline

Let a family of experiments be indexed by the set $\mathbb{N} = \{1, 2, \dots\}$. Each experiment yields an outcome in the set S (technical details are suppressed in this section). Thus $\Omega = S^\infty$ is the set of all possible sample paths. A probability measure P on Ω is *exchangeable* if

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

for all finite permutations π of \mathbb{N} . De Finetti [16] shows that exchangeability is equivalent to the following representation: There exists a (necessarily unique) probability measure μ on $\Delta(S)$ such that

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

where, for any probability measure ℓ on S (written $\ell \in \Delta(S)$), ℓ^∞ denotes the corresponding i.i.d. product measure on Ω . Thus beliefs are i.i.d. conditional on the unknown parameter ℓ ; learning is then modeled by Bayesian updating of beliefs about the parameter. Kreps [26, Ch. 11] refers to de Finetti’s celebrated result as “the fundamental theorem of (most) statistics” because of the justification it provides for the analyst to view samples as being independent and identically distributed with unknown distribution function; and he argues for the importance of exchangeability and de Finetti’s Theorem as normative decision tools.

Though the de Finetti Theorem can be viewed as a result in probability theory alone, it is typically understood in economics as a prescription for imposing structure on the predictive prior P in the subjective expected utility model of choice. That is also how we view it and accordingly we provide a decision-theoretic generalization of de Finetti’s result that we view as largely normative. Specifically, we consider preference on a domain of (Anscombe–Aumann) acts that conforms to Schmeidler’s [35]. Choquet expected utility where the capacity is a belief function—we call this model *belief function utility*.¹ Using the latter as the basic framework, we then impose two simple axioms—Exchangeability (the preference counterpart of de Finetti’s assumption) and Weak Orthogonal Independence (relaxing the Independence axiom). These axioms are shown (Theorem 3.1) to characterize the following representation for the belief function κ on Ω (see the noted theorem for the corresponding representation of utility):

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

where $Bel(S)$ denotes the set of all belief functions on S , μ is a (necessarily unique) probability measure on $Bel(S)$, and ν^∞ denotes a suitable “i.i.d. product” of the belief function ν . The de Finetti–Savage model is the special case where (the Independence axiom is satisfied and hence) each ν in the support of μ is additive.

¹ Belief functions are special cases of capacities (or “non-additive probabilities”), sometimes referred to as totally, completely, or infinitely monotone capacities. They originated in Dempster [8]; definitions for more general settings can be found, for example, in Philippe, Debs and Jafray [34], and Molchanov [33]. See Appendix A for details on belief functions and the corresponding utility functions.

The model has several attractive features. First, it accommodates ambiguity aversion, and thus Ellsberg-style behavior, through the generalization from expected utility to belief function utility. The second point is specific to a setting with repeated experiments: the representation (1.1) suggests uncertainty about the probability law describing each experiment but also certainty that experimental outcomes are identically and independently distributed, thus ruling out any role for imprecise information about how the experiments may differ and/or be related (a critique that is made behavioral below). In contrast, (1.2) relaxes the noted certainty because each ν^∞ is associated with its core, a nonsingleton set of measures on S^∞ , and the core contains many measures that are either not identical products or not product measures at all, thus permitting ambiguity about heterogeneity and correlation to matter.

The de Finetti representation is important in part because it provides formal justification for the reference to ‘parameters,’ features that are believed to be common across all experiments and that can (to some degree) be learned. Our representation theorem characterizing (1.2) provides foundations for generalizing the notion of parameter to a belief function ν over each experiment. In the Bayesian framework, knowledge of the parameter determines a unique probability law over the sequence of experiments, and thus equips the individual with all she needs to predict outcomes and to choose between acts. In contrast, in our model knowledge of the parameter ν determines only a nonsingleton set (the core of ν^∞) of probability laws over sequences of experiments. Because of the individual’s inability or unwillingness to make a probabilistic prediction even given knowledge of her parameter, we refer to her as having an *incomplete theory* of her environment. Finally, in the Bayesian model and under common assumptions, the individual is certain that she will learn the true parameter asymptotically given enough data. Accordingly, the model leaves no room for doubt about what is well-understood or for more modest ambitions about what is learnable. Our generalization accommodates less extreme self-confidence (or naivete); learning in our model is taken to be the modified form of Bayesian updating described in Epstein and Seo [15] and outlined below.

While we generalize de Finetti in the ways just noted, we share with his model the restriction imposed by exchangeability, namely that experiments are indistinguishable in the sense that, for example, betting on a subset of outcomes in experiment i is indifferent to the corresponding bet on experiment j . In the probability framework, this symmetry has been relaxed to “partial symmetry” or “partial exchangeability” (de Finetti [17], Link [27], Diaconis and Freedman [10]), where the set of experiments can be partitioned into a finite number of classes such that exchangeability holds within each class but not globally. We leave to future work a corresponding generalization of our model.

There is another ‘epistemic’ perspective on the role of parameters and the source of theory incompleteness in our model. We borrow the well-known Dempster [8] and Shafer [36] intuition for belief functions applied to beliefs about Ω . They postulate imperfect understanding of Ω modeled via an auxiliary epistemic state space $\widehat{\Omega}$ and a correspondence $\Gamma : \widehat{\Omega} \rightsquigarrow \Omega$. Assuming that beliefs on $\widehat{\Omega}$ can be represented by a probability measure, denoted M , and given a conservative attitude, the triple $(\widehat{\Omega}, M, \Gamma)$ determines (and represents) a belief function κ via

$$\kappa(A) = M(\{\widehat{\omega} \in \widehat{\Omega} : \Gamma(\widehat{\omega}) \subset A\}), \quad A \subset \Omega. \quad (1.3)$$

To interpret κ , think of the case where $\widehat{\Omega}$ is a finite partition of Ω (each auxiliary state $\widehat{\omega}$ is a cell of that partition), and $\Gamma(\widehat{\omega}) = \widehat{\omega}$. Then, on the payoff relevant space Ω , the individual is able to assign the sharp probability $M(\widehat{\omega})$ to each subset $\Gamma(\widehat{\omega})$. However, she is completely ignorant of likelihoods within each $\Gamma(\widehat{\omega})$. This “explains” why, when forming beliefs about a subset $A \subset \Omega$, she does not assign any weight to auxiliary states $\widehat{\omega}$ for which $\Gamma(\widehat{\omega}) \setminus A$ is nonempty—even if this

set is “small” in some sense, ignorance within $\Gamma(\widehat{\omega})$ and a conservative attitude prevent her from taking a stand and evaluating $\Gamma(\widehat{\omega}) \setminus A$ as being highly unlikely. More generally, the belief function κ models an individual who is extremely confident about likelihoods for some aspects of her environment (and thus beliefs are probabilistic for these) and is completely ignorant of likelihoods for other aspects (where any conditional probability law might apply and no weighting of alternative laws is possible). In other words and roughly speaking, belief functions generalize probability measures by permitting some ignorance.²

Consider next belief functions in our setting with repeated experiments. Note that though the preceding refers to the state space Ω of all infinite sequences of outcomes for all experiments, the Cartesian product structure for Ω is not reflected— Ω could be any abstract state space (for example, the identical constructs apply equally when Ω is replaced by S and thus the reference is to beliefs about a single experiment). In particular, κ does not in general admit the representation (1.2), and the triple $(\widehat{\Omega}, M, \Gamma)$ does not identify parameters—factors common to all experiments.

We consider instead the following tuple $(\widehat{S}^\infty, m^\infty, G^\infty, \Theta, \mu)$, which we call a *theory*, where: $\widehat{S}_i = \widehat{S}$ is an auxiliary state space used when forming beliefs about S_i and $\widehat{S}^\infty = \widehat{S}_1 \times \widehat{S}_2 \times \dots$ is the infinite Cartesian product; Θ is a set of parameters; for each θ in Θ , $G(\cdot | \theta)$ is a correspondence from \widehat{S} into S , and $G^\infty(\cdot | \theta) : \widehat{S}^\infty \rightsquigarrow \Omega$ is the correspondence given by

$$G^\infty((\widehat{s}_1, \dots, \widehat{s}_i, \dots) | \theta) = \prod_{i=1}^\infty G(\widehat{s}_i | \theta); \tag{1.4}$$

m^∞ denotes the collection $\{m_\theta^\infty : \theta \in \Theta\}$ of measures on \widehat{S}^∞ , where, for each θ , m_θ^∞ is the i.i.d. product of $m_\theta \in \Delta(\widehat{S})$; and μ is the prior over Θ describing beliefs about the true parameter value. Thus the preceding specializes part of the Dempster–Shafer triple $(\widehat{\Omega} = \widehat{S}^\infty$, and other product structures are assumed), and also extends their triple primarily by adding the set Θ of parameters and the prior μ . The interpretation is that the outcomes of experiment i are assumed by the individual to depend on the realization of \widehat{s}_i in \widehat{S} and on the unknown parameter θ ; \widehat{s}_i is unobserved but is thought to be distributed according to the probability law m_θ . However, the theory is incomplete in that θ and \widehat{s}_i determine only the set of outcomes $G(\widehat{s}_i | \theta)$, and the theory is silent on how the realized outcome s_i is selected from $G(\widehat{s}_i | \theta)$. With regard to the perception of the entire sequence of experiments, θ and $G(\cdot | \theta)$ are assumed to be common to all experiments, and the \widehat{s}_i s are taken to be i.i.d. according to m_θ . The final component is $G^\infty(\cdot | \theta)$ in (1.4). The Cartesian product structure implies that if for each i , s_i is a conceivable outcome in the i th experiment (in isolation) given \widehat{s}_i (and θ), then (s_1, s_2, \dots) is a conceivable sequence given $(\widehat{s}_1, \widehat{s}_2, \dots)$, which expresses complete ignorance about the relationship between selection mechanisms across experiments. This is a consequence of ignorance of the selection mechanism, which implies that there is no basis for understanding how selection, and thus realized outcomes, may differ or be related across experiments.

A theory induces belief functions as in (1.3). For each θ , define ν_θ^∞ by

$$\nu_\theta^\infty(A) = m_\theta^\infty(\{(\widehat{s}_1, \dots, \widehat{s}_i, \dots) \in \widehat{S}^\infty : \prod_{i=1}^\infty G(\widehat{s}_i | \theta) \subset A\}), \quad A \subset \Omega. \tag{1.5}$$

The theory is *incomplete* when $\nu_\theta^\infty(\cdot)$ is nonadditive for a nonnegligible set of parameters θ , which is the case when $G(\cdot | \theta)$ is multi-valued sufficiently often. A theory *represents* the Choquet expected utility preference having belief function κ given by

² See Wakker [41].

$$\kappa(\cdot) = \int_{\Theta} v_{\theta}^{\infty}(\cdot) d\mu(\theta). \tag{1.6}$$

Our axioms also characterize the set of preferences that can be represented in this way by some theory. In addition, we show in Section 3.2 that while the representation (1.2) is a special case (having $\Theta = Bel(S)$, for example), it is also canonical in that any capacity κ defined as in (1.6) admits representation also as in (1.2).

The paper proceeds as follows. The introduction concludes with two (related) running examples; the first concerns Ellsberg-style urns, and the second concerns entry games with multiple Nash equilibria such as have been studied in the applied IO literature. Axioms and the implied representation of utility are described in Sections 2 and 3 respectively. Though these concern ex ante preference only, we outline how by applying Epstein and Seo [15] the model can be extended to include sampling and updating. To buttress the normative case for our model, Section 3.4 describes a connection between prior beliefs about parameters and about empirical frequencies that may aid in calibrating the former; this result exploits a law of large numbers (LLN) for belief functions due to Maccheroni and Marinacci [28]. Two applications are presented in Section 4. The first considers the classical problem of the optimal prediction of empirical frequencies; and the second shows that ambiguity can limit participation in equity markets even given the possibility of diversification, thus extending Dow and Werlang [11]. Section 5 concludes with a discussion of related literature and an explanation of the value-added herein. Proofs and technical details are collected in Appendices A–E.

1.2. Running examples: urns and entry games

Consider a decision-maker facing an infinite sequence of Ellsberg urns. One ball will be drawn from each urn with all draws being simultaneous. The decision-maker must choose between bets on the outcomes of the sequence of draws. She is told only that each contains 100 balls that are either red (R) or blue (B). Thus she is not given any reason to be certain that the compositions are identical, nor to be confident that the urn compositions are unrelated or related in any particular way; hence she may wish to choose bets that are robust to this uncertainty. As shown in the sequel, our model can accommodate such robustness.

Each experiment is a draw from an urn with possible outcomes R and B ; thus $S = \{R, B\}$. Belief functions on S , and for any binary state space, are particularly simple—they are in one-to-one correspondence with *probability intervals* for drawing red. That is, any belief function ν on S induces the probability interval $[\nu(R), 1 - \nu(B)]$. Conversely, given $\theta = (\theta_1, \theta_2)$ and the corresponding interval $[\theta_1, \theta_2] \subset [0, 1]$, then $[\theta_1, \theta_2] = [\nu_{\theta}(R), 1 - \nu_{\theta}(B)]$ for ν_{θ} represented by the Dempster–Shafer triple $(\widehat{S}, m_{\theta}, G(\cdot | \theta))$, where

$$\widehat{S} = \{\{R\}, \{B\}, \{R, B\}\}, \tag{1.7}$$

$$m_{\theta}(\{R\}) = \theta_1, \quad m_{\theta}(\{B\}) = 1 - \theta_2, \quad m_{\theta}(\{R, B\}) = \theta_2 - \theta_1, \tag{1.8}$$

and

$$G(\{R\} | \theta) = R, \quad G(\{B\} | \theta) = B, \quad G(\{R, B\} | \theta) = S. \tag{1.9}$$

Probability intervals can arise in the mind of the decision-maker if, for example, she adopts the following view of how the urns are constructed. She hypothesizes that the fraction λ of the 100 balls is selected once and for all by a single administrator and then placed in each urn, while the other $(1 - \lambda)$ 100 vary across urns in a way that is not understood and about which she is

completely agnostic. If ρ denotes the proportion of red in the common group of balls, then the probability of drawing red from any urn lies between $\lambda\rho$ and $\lambda\rho + (1 - \lambda)$. Thus the unknown parameter that is common across urns can be thought of as the probability interval for red given by $J = [\lambda\rho, \lambda\rho + (1 - \lambda)]$, or equivalently as the minimum and maximum probabilities $\lambda\rho$ and $\lambda\rho + (1 - \lambda)$ respectively.

Other components of our model of this choice situation will be illustrated as we go through a more detailed description of the model.

1.3. An entry game

We take the urns setting as a canonical example of choice where uncertainty is derived from many ‘exchangeable’ random events. Here we describe a more concrete choice situation that has played a role in the applied IO literature and that is essentially isomorphic to the urns setting.

Consider a policy maker (PM) who must choose a policy that pertains to a number of markets, in each of which there are two (potential) firms. The consequences of the policy depend on firm entry decisions which are uncertain (hence policies are acts). In the i th market, firms $j = 1, 2$ play the entry game described by the payoff matrix shown:

	out	in
out	0, 0	0, $-\widehat{s}_{i2}$
in	$-\widehat{s}_{i1}, 0$	$\eta^{1/2} - \widehat{s}_{i1}, \eta^{1/2} - \widehat{s}_{i2}$

The parameter η lies in $[0, 1]$ and the \widehat{s}_{ij} ’s are observed by players but not by the PM. She views η as fixed and common across markets and the \widehat{s}_{ij} ’s as uniformly distributed on $[0, 1]^2$ for each i and i.i.d. across markets. The PM’s theory is that the outcome in each market is a pure strategy Nash equilibrium. However, her theory is incomplete because she does not understand equilibrium selection at all and this is important because there may be multiple equilibria: the set of Nash equilibria in market i is given by

$$\begin{array}{|c|c|}
 \hline
 \{T, N\} & \text{if } 0 \leq \widehat{s}_{i1}, \widehat{s}_{i2} \leq \eta^{1/2} \\
 \hline
 \{N\} & \text{otherwise,} \\
 \hline
 \end{array} \tag{1.10}$$

where T denotes the outcome where two firms enter and N the outcome where none enter. Thus, even given knowledge of η , without taking a stand on selection PM can be sure only that the probability of T lies in the interval $[0, \eta]$.

This entry game, taken from Jovanovic [23], serves as a canonical example in a literature on the empirical analysis of complete information entry games with multiple equilibria; see, for example, Tamer [40,39], Ciliberto and Tamer [7] and the references therein. For our purposes, many of the special features of the above game are made purely for simplicity, including: two firms in each market, the particular functional forms for payoffs, and the uniform distribution for the unobserved heterogeneity represented by the \widehat{s} ’s. In fact, our model accommodates any setting where the decision-maker’s theory of her environment can be represented as described above surrounding (1.5). As explained shortly, the restriction to pure strategy equilibria is important for our approach to modeling policy choice.

To elaborate on agnosticism about selection, imagine the PM having the following perspective. She believes that a complete theory of equilibrium selection exists in principle, and that selection could be explained and predicted given a suitable set of explanatory variables, but she (and most economists) cannot identify these “omitted variables.” As a result, not only can she not

assign a probability to T being selected in any given market, neither does she understand how selection may differ or be correlated across markets. Thus she seeks to make decisions that are robust to heterogeneity and correlation of an unknown form.

A “reduced form” of this choice situation is almost identical to the urns setting with the change in ‘colors’ from R and B to T and N . In particular, in both cases, there is the perception that two factors underlie experimental outcomes: one that is fixed across all experiments, (the fixed portion of each urn above and the parameter η here), and the second that is idiosyncratic and not understood, (the remaining portion of each urn above and selection here). In both cases, probability intervals arise as descriptions of the factor common to all experiments; here the interval $[0, \eta]$ for the outcome T is in one-to-one correspondence with the parameter η . A small difference is that the two ‘colors’ T and N do not appear symmetrically because N can be realized also as a unique equilibrium—thus one would expect a strict preference to bet on N over T in any market.

A Dempster–Shafer triple $(\widehat{S}, m_\eta, G(\cdot | \eta))$ arises naturally in the entry game context: take m_η to be the uniform distribution on $\widehat{S} = [0, 1]^2$, and let $G(\cdot | \eta)$ be the Nash equilibrium correspondence:

$$G(\widehat{s}_{i1}, \widehat{s}_{i2} | \eta) = \begin{cases} \{T, N\} & \text{if } 0 \leq \widehat{s}_{i1}, \widehat{s}_{i2} \leq \eta^{1/2} \\ \{N\} & \text{otherwise.} \end{cases} \tag{1.11}$$

The triple induces a belief function v_η on S which is in one-to-one correspondence with the parameter η . The restriction to pure strategy equilibria is evident here—with mixed strategies, the equilibrium correspondence $G(\cdot | \eta)$ would map into subsets of $\Delta(\{in, out\}^2)$, and v_η would be a belief function over the latter simplex but not over S .

2. Foundations

Consider a sequence of experiments, each of which yields an outcome in the compact metric space S ; we refer back often to the urns example where $S = \{R, B\}$. The payoff to any chosen physical action depends on the realized state in the state space Ω given by

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

Objects of choice are (Borel measurable and simple, that is, finite-ranged) acts $f : \Omega \rightarrow [0, 1]$. The set of all acts is \mathcal{F} . Binary acts are called bets. The bet $\mathbf{1}_A$ that the event $A \subset \Omega$ will occur is denoted simply A . For example, $R_1 B_2$ is the bet that pays 1 if red is drawn from the first urn and blue from the second, and the bet that the first two urns yield the same color is $\{R_1 R_2, B_1 B_2\}$.

Payoffs to acts should be interpreted as measured in utils, which are derived from an expected utility ranking of objective lotteries. Denominating payoffs in utils can be justified via a more primitive Anscombe and Aumann [3] formulation of choice under uncertainty. Because these details are standard, we simplify and adopt the reduced form above. Note that with payoffs denominated in utils, and given a vNM ranking of objective lotteries, one can view the individual *as though* she were risk neutral.

We study *ex ante* preference \succeq over acts; in the next section we outline how the model can be extended to include also the updating of preference after observing the outcomes of finitely many experiments.

To state the first axiom, we must define “belief function utility”. Henceforth refer to $(\widehat{\Omega}, M, \Gamma)$ as a *Dempster–Shafer triple* (for Ω) if it is defined as in the introduction and if also: $\widehat{\Omega}$ is compact metric, M is a Borel probability measure, and the correspondence Γ is weakly measurable and

nonempty-compact-valued.³ Any function on the Borel σ -algebra of Ω , (which can be taken to be S^∞ as above, or any other compact metric space), that can be constructed via a Dempster–Shafer triple $(\widehat{\Omega}, M, \Gamma)$ as in (1.3) is called a *belief function* on Ω .⁴ A function $U : \mathcal{F} \rightarrow \mathbb{R}$ is called a *belief function utility* (for the state space Ω) if there exists a belief function κ on Ω such that

$$U(f) = \int_{\Omega} f d\kappa, \text{ for all } f \text{ in } \mathcal{F}. \tag{2.1}$$

Here integration is in the sense of Choquet and thus every belief function utility is a special case of Choquet expected utility (Schmeidler [35]). Say that $(\widehat{\Omega}, M, \Gamma)$ represents the belief function κ and also the utility function U and the corresponding preference \succeq . Note that $U(f)$ is the *certainty equivalent* of f —that payoff which if received in every state would be indifferent to f .

We adopt three axioms for the preference \succeq .

Axiom 1 (Belief Function Utility). \succeq admits representation by a belief function utility.

This axiom is not completely satisfactory because it is not stated in terms of behavior which is presumably the only observable. However, Epstein, Marinacci and Seo [12] and Gul and Pendorfer [20] describe behavioral foundations for (2.1), albeit in somewhat different formal frameworks. Because modeling ambiguity aversion in the abstract is not our focus, we move on to study the special features arising from the presence of repeated experiments. There is a parallel with de Finetti, who took subjective expected utility (or at least a predictive prior) as given and explored the implications of exchangeability for a setting with repeated experiments. We take belief function utility as given and focus on additional structure that is of interest given repeated experiments. The next two axioms describe the individual’s perception of experiments and how they are related.

Given subjective expected utility preferences, de Finetti’s assumption that the prior is exchangeable is equivalent to the following restriction on preference. Let Π be the set of (finite) permutations on \mathbb{N} . For $\pi \in \Pi$ and $\omega = (s_1, s_2, \dots) \in \Omega$, let $\pi\omega = (s_{\pi(1)}, s_{\pi(2)}, \dots)$. Given an act f , define the permuted act πf by $(\pi f)(s_1, \dots, s_n, \dots) = f(s_{\pi(1)}, \dots, s_{\pi(n)}, \dots)$. For example, if $f = R_1 B_2$ and π switches 1 and 2, then $\pi f = B_1 R_2$. An act is said to be *finitely-based* if it depends on the outcomes of only finitely many experiments. Thus the bet A is finitely-based if as an event, A restricts the outcomes of only finitely many experiments.

Axiom 2 (Exchangeability). For all finitely-based bets A and permutations π ,

$$A \sim \pi A.$$

Exchangeability is intuitive when information about the experiments is symmetric, and thus you are not given any reason to distinguish between them. Note, however, that symmetry of

³ Throughout every compact metric space is endowed with the induced Borel σ -algebra and probability measures are understood to be Borel measures.

A correspondence $\Gamma : \widehat{\Omega} \rightsquigarrow \Omega$, where Ω is metric, is weakly measurable if $\{\widehat{\omega} : \Gamma(\widehat{\omega}) \subset A\}$ is a (Borel) measurable subset of $\widehat{\Omega}$ for every closed $A \subset \Omega$. If Γ is compact-valued, as here, then weak measurability is equivalent to the property that $\{\widehat{\omega} : \Gamma(\widehat{\omega}) \subset A\}$ is measurable for every open $A \subset \Omega$ (Aliprantis and Border [1], Lemma 18.2).

⁴ Equivalent definitions of belief functions are described in Appendix A. If Γ is singleton-valued and hence a random variable, then κ is a probability measure and (1.3) is the familiar formula for computing induced distributions.

information does not imply that information is substantial; in fact, as in the urns example, there could be very little information available at all about any of the experiments and about how they are related. Thus Exchangeability is entirely consistent with ambiguity about both correlation and heterogeneity. The next axiom leaves room for such ambiguity. It does so by suitably relaxing the Independence axiom to permit randomization to have positive value in some circumstances.

Refer to bets A' and A as mutually *orthogonal* if they depend on disjoint sets of experiments; write $A' \perp A$. (Note that a bet on Ω , which is constant act, is orthogonal to every bet A .) Our main axiom is⁵:

Axiom 3 (*Weak Orthogonal Independence (WOI)*). For all finitely based bets A' , A and B such that $A' \perp B$ and $A \perp B$,

$$A' \succeq A \iff \frac{1}{2}A' + \frac{1}{2}B \succeq \frac{1}{2}A + \frac{1}{2}B.$$

The Independence axiom requires the similar invariance of rankings for all (not necessarily orthogonal or binary) acts. We argue that Independence is too strong given a concern with unknown correlation and heterogeneity. In fact, one can illustrate behaviorally three separate kinds of ambiguity that seem relevant to the urns example and that are excluded by Independence but permitted by WOI. The first is simply ambiguity about the outcome in any *single* experiment, which is illustrated by the following Ellsberg-style behavior contradicting Independence:

$$\frac{1}{2}R_1 + \frac{1}{2}B_1 \succ R_1 \sim B_1. \quad (2.2)$$

Gilboa and Schmeidler [19] describe the value of such randomization as due to its smoothing out ambiguity (indeed, the mixed bet yields the payoff $\frac{1}{2}$ with certainty) or, adapting finance terminology, because the bets being mixed “hedge” one another.

The other two kinds of ambiguity have to do with how the compositions of *different* urns are related. Consider the ranking

$$\frac{1}{2}R_1 + \frac{1}{2}B_1 \succ \frac{1}{2}R_1 + \frac{1}{2}B_2. \quad (2.3)$$

The act on the left perfectly hedges uncertainty about the first urn and yields $\frac{1}{2}$ with certainty. But the act on the right also involves uncertainty about possible differences between urns. For example, if the first urn is biased towards red and the second is biased towards blue, that is a good scenario for $\frac{1}{2}R_1 + \frac{1}{2}B_2$. However, under the reverse scenario, the latter act is unattractive. Thus if both scenarios are considered possible, and there is aversion to uncertainty about which is true, then the indicated ranking follows. In this way, aversion to ambiguous heterogeneity suggests (2.3).

The following behavior, which is consistent with WOI but not with Independence, reveals a concern with correlation which we take to mean roughly a concern that the urns' compositions may follow some unknown “patterns.” Consider betting that the colors drawn from the first two urns are identical versus betting that they are identical in the first and third urns. Exchangeability implies indifference. However, there is intuition, once again based on smoothing out uncertainty, for the following rankings:

⁵ Talagrand [38] studies belief functions ν that satisfy $\nu(A) = \nu(\pi A)$, for all A and π , and shows that this symmetry property alone implies very little structure. This supports our view that the next axiom is our main assumption.

$$\begin{aligned} & \frac{1}{2}\{R_1 R_2, B_1 B_2\} + \frac{1}{2}\{R_1 R_3, B_1 B_3\} \\ & \succ \{R_1 R_2, B_1 B_2\} \sim \{R_1 R_3, B_1 B_3\}. \end{aligned} \quad (2.4)$$

Finally, we turn to behavior that is excluded by WOI. The following violation of WOI reveals something about the nature of the robustness that it permits. Suppose the individual is certain that either all urns contain only red balls or that they all contain only blue balls, though she is unsure which is true. Then, if she is averse to ambiguity about which of these two possibilities is valid, one would expect that

$$\frac{1}{2}R_1 + \frac{1}{2}B_2 \succ R_1 \sim B_2. \quad (2.5)$$

Such a value for randomization contradicts WOI, at least given also Belief Function Utility (see Lemma B.2). Since this scenario is one of “perfect correlation,” one may wonder whether or in what sense our model indeed accommodates aversion to ambiguity about correlation. To be perfectly clear, the axiom and model admit perfect correlation as a possibility in the mind of the individual. It is the simultaneous exclusion of all other correlation patterns on her part, or certainty that the urns are perfectly correlated, that is contradicted by the axiom, given also aversion to ambiguity about correlation.⁶ For instance, the strict preference indicated in (2.5) is not intuitive if the individual admits also the possibility that the first urn being biased towards red (blue) makes it highly likely that the second is biased towards blue (red). The bottom line is that, as will be apparent from the representation, we model an individual who is seeking robustness against dependence of unknown form between urn compositions because she has no basis for excluding any particular patterns a priori.

To understand better what else the axiom excludes, recall that the Gilboa–Schmeidler intuition is that hedging two bets can be valuable when there is a common ambiguous factor underlying them. For greater clarity, consider the entry game. Then the parameter η is common to all markets. Therefore, if there is ambiguity about η , then randomization between indifferent bets on different markets can be valuable, which as above contradicts WOI and Belief Function Utility. In other words, ambiguity about parameters is excluded, which explains why in (1.2) and (1.6) there is a single prior μ representing uncertainty about parameters. The other factor influencing outcomes in each market is the selection mechanism, which we take to be poorly understood or ambiguous. Therefore, we interpret the absence of hedging gains as expressed in WOI to mean that the selection mechanism is not common, or more accurately, that because selection is so poorly understood, there is no basis for believing that the mechanisms in two different markets are connected or related in any particular way. Again, this feature is reflected in the implied representation of preference, most explicitly in the Cartesian product structure of any representing theory (as in the discussion of (1.4)).

3. Representation

3.1. The main result

To state our main result, we need some preliminaries on belief functions. Let the Dempster–Shafer triple (\widehat{S}, m, G) represent the belief function ν on S in the sense introduced above; recall

⁶ If she is indifferent in (2.5), then the ranking is consistent with WOI; indeed, it is consistent not only with our model but also with the exchangeable Bayesian model (1.1); for instance, let the prior μ assign positive probability only to the measures ℓ that are degenerate at R or at B .

that every belief function on S can be represented in this way. Denote by $Bel(S)$ the set of all belief functions on S .⁷ Each ν in $Bel(S)$ can be thought of as describing beliefs about a single experiment.

Given any belief function ν on S and a representing triple (\widehat{S}, m, G) , denote by ν^∞ the belief function on Ω that is represented by the triple $(\widehat{S}^\infty, m^\infty, G^\infty)$, where m^∞ is the ordinary i.i.d. product of the probability measure m , and G^∞ is the correspondence $G^\infty : (\widehat{S})^\infty \rightsquigarrow \Omega = S^\infty$ given by

$$G^\infty(\widehat{s}_1, \widehat{s}_2, \dots) = G(\widehat{s}_1) \times G(\widehat{s}_2) \times \dots \tag{3.1}$$

We refer to ν^∞ as the “i.i.d. product” of ν .⁸ When G is singleton-valued, then ν is a probability measure and ν^∞ is the usual i.i.d. product for measures.

We can now state our main result.

Theorem 3.1. *Let \succeq be a preference order on the set of acts \mathcal{F} . Then the statements (a) and (b) are equivalent:*

(a) \succeq satisfies Belief Function Utility, Exchangeability and Weak Orthogonal Independence.

(b) There exists a (necessarily unique) Borel probability measure μ on $Bel(S)$ such that \succeq has a utility function U of the form

$$U(f) = \int_{\Omega} f d\kappa, \text{ for every } f \text{ in } \mathcal{F}, \tag{3.2}$$

where κ is the belief function given by

$$\kappa(A) = \int_{Bel(S)} \nu^\infty(A) d\mu(\nu), \text{ for every Borel } A \subset \Omega. \tag{3.3}$$

With regard to sufficiency of the axioms, the Choquet expected utility representation in (3.2) is simply a restatement of the assumption Belief Function Utility. The main content of the theorem is in the “conditionally i.i.d.” form derived for κ , paralleling and generalizing de Finetti’s classic form in (1.1); the latter is obtained if, for example, WOI is strengthened to the usual Independence axiom. In common with the classic model, the prior μ is unique and it completely determines the preference order.

The representation suggests that ambiguity about heterogeneity and correlation are accommodated. To see why, note first that $U(\cdot)$ is an average (using μ) of utility functions of the form

$$V_\nu(f) \equiv \int f d\nu^\infty; \tag{3.4}$$

thus it suffices to consider V_ν , which is based on the i.i.d. product belief function ν^∞ . As outlined in Appendix A, every belief function utility conforms with the maxmin model (Gilboa and Schmeidler [19]) with set of predictive priors equal to the core of the belief function. For V_ν we have

⁷ Endow $Bel(S)$ with the topology for which $\nu_n \rightarrow \nu$ if and only if $\int f d\nu_n \rightarrow \int f d\nu$ for every continuous function f on S , where the integral is in the sense of Choquet. Then $Bel(S)$ is compact metric.

⁸ Appendix A shows that ν^∞ is well-defined: if (\widehat{S}, m, G) and (\widehat{S}', m', G') both represent ν , then they both lead to the same belief function on S^∞ .

$$V_\nu(f) = \inf_{P \in \text{core}(\nu^\infty)} \int_{\Omega} f dP,$$

where⁹

$$\begin{aligned} \text{core}(\nu^\infty) &\equiv \{P \in \Delta(\Omega) : P(\cdot) \geq \nu^\infty(\cdot)\} \\ &= \int_{\hat{S}^\infty} \Delta(G(\hat{s}_1) \times G(\hat{s}_2) \times \dots) dm^\infty(\hat{s}_1, \hat{s}_2, \dots). \end{aligned}$$

The point to note is that the indicated simplex includes *all* joint distributions over $\prod_{i=1}^\infty G(\hat{s}_i)$; thus the core contains both nonidentical product measures and nonproduct measures.

At the more meaningful behavioral level, the model accommodates the rankings (2.3)–(2.4) for bets on the urns that we have interpreted in terms of aversion to ambiguity about heterogeneity and correlation respectively.¹⁰ In fact, these rankings are accommodated even in the special case where μ assigns probability 1 to a particular belief function ν , where

$$\nu(R), \nu(B) > 0, \nu(R) + \nu(B) < 1.$$

Then, for the first ranking,

$$\begin{aligned} V_\nu\left(\frac{1}{2}R_1 + \frac{1}{2}B_2\right) &= \frac{1}{2}V_\nu(R_1) + \frac{1}{2}V_\nu(B_2) \\ &= \frac{1}{2}\nu(R) + \frac{1}{2}\nu(B) < \frac{1}{2} = V_\nu\left(\frac{1}{2}R_1 + \frac{1}{2}B_1\right), \end{aligned}$$

where the first equality follows from Lemma B.2. For (2.4), abbreviate the bet $\{R_1 R_2, B_1 B_2\}$ by f and let π be the permutation that switches the second and third markets. Then, (see supporting details at the end of Appendix A),

$$\begin{aligned} V_\nu(f) &= (\nu(R))^2 + (\nu(B))^2 \\ &< (\nu(R))^2 + (\nu(B))^2 + \nu(R)\nu(B)(1 - \nu(R) - \nu(B)) \\ &= V_\nu\left(\frac{1}{2}f + \frac{1}{2}\pi f\right). \end{aligned} \tag{3.5}$$

Remark 1. While referring to ambiguity about correlation, we have also referred to ν^∞ as an “i.i.d. product” of ν and to (3.3) as a “conditionally i.i.d.” representation. The latter terminology is used partly because ν^∞ is the usual i.i.d. product when ν is a measure. It is suggested also by the construction of ν^∞ , and by the fact that ν^∞ satisfies a product property for rectangles; for example, $\nu^\infty(A_1 \times A_2 \times S^\infty) = \nu^\infty(A_1 \times S \times S^\infty) \cdot \nu^\infty(S \times A_2 \times S^\infty)$ for all $A_1 \subset S_1$ and $A_2 \subset S_2$, and similarly for all other rectangles. Though we would like to be more precise about the meaning of “stochastic independence,” it is well-known that the latter is multifaceted if there is ambiguity and that it is not well-understood behaviorally (Ghirardato [18]). One contribution made herein to the surrounding literature can be noted. It is known that, because ν is not additive, there is more than one way to extend ν to a belief function on S^∞ that satisfies the noted property on rectangles, but the differing implications for behavior are not understood. In the case of finitely many experiments, the rule used here for forming the i.i.d. product ν^∞ is proposed by

⁹ The integral that follows is an Aumann integral and the characterization of the core given is based on Philippe et al. [34, Thm. 3].

¹⁰ Of course, it also accommodates the standard Ellsberg-style behavior in (2.2).

Dempster [8,9] and studied by Hendon et al. [21] and Ghirardato [18]. The latter shows that it is the only product rule for belief functions such that the product (i) is also a belief function, and (ii) satisfies a mathematical property called the Fubini property. A contribution of our model is that the product rule $\nu \mapsto \nu^\infty$ emerges as an implication of assumptions about preference.

3.2. Theory incompleteness and parameters

Besides ambiguity about heterogeneity and correlation, another motivation for the model is to provide foundations for a more general notion of “parameter.” De Finetti’s conditionally i.i.d. representation (1.1) for subjective predictive priors is commonly interpreted in terms of uncertainty about the parameter ℓ , the probability law that describes every experiment; and the Savage/Anscombe–Aumann axioms plus Exchangeability provide behavioral foundations. Viewing the model as a normative model, we take the preceding as a description of how the individual perceives and theorizes about her environment; one can also view it in an “as if” vein. In the same way, our representation (3.3) suggests the interpretation whereby the individual characterizes every experiment by the same (though uncertain) belief function ν ; and our weaker axioms provide corresponding foundations for taking $Bel(S)$ rather than $\Delta(S)$ as the parameter space. In the de Finetti model, given the inherently stochastic nature of each experiment, knowledge of the parameter provides the individual with all she needs to know to predict outcomes of experiments and hence to choose between acts. In contrast, certainty that the belief function parameter is ν determines only the set of probability laws *core* (ν^∞). Because knowledge of the parameter does not permit a unique probabilistic prediction, we refer to the individual’s theory as incomplete.

The preceding takes a more concrete form in the urns example. Then, (and similarly for any binary S), each belief function ν can be identified with the probability interval $[\nu(R), 1 - \nu(B)]$ for a red draw. Thus instead of each urn being characterized by a common single number in the unit interval, as in the Bayesian model, here each urn is characterized by a common interval. The interval is nondegenerate in general in order to model the individual’s concern with idiosyncratic and poorly understood differences between urns, and thus a lack of confidence that the urns have identical compositions. The idea is that though the interval is common, and thus urns are indistinguishable—for example, there is indifference between betting on red from urn i and the corresponding bet on urn j —it is understood that any probability in the interval can apply to urn i and that any other probability in the interval may apply to urn j . Accordingly, even given knowledge of ν , or equivalently, knowledge of the probability interval for red, there remains ambiguity about urn compositions and about how they are related. (In the entry game variant of the urns example as described in Section 1.2, knowledge of $\nu = \nu_\eta$ amounts to knowledge of η , but there remains ambiguity about selection.)

Next we formalize and generalize some of the preceding by means of tuples $(\widehat{S}^\infty, m^\infty, G^\infty, \Theta, \mu)$ of the form described in the introduction. Each tuple satisfying the conditions below is referred to as a *theory*. In particular, Θ is a (compact metric) parameter space, uncertainty about the identity of the true parameter is represented by μ , a probability measure on Θ , m^∞ denotes the collection $\{m_\theta^\infty : \theta \in \Theta\}$ of measures on \widehat{S}^∞ , and $G^\infty(\cdot | \theta)$ is a correspondence from \widehat{S}^∞ to Ω having the Cartesian product structure in (1.4). Require that: (i) for each θ , $(\widehat{S}^\infty, m_\theta^\infty, G^\infty(\cdot | \theta))$ is a Dempster–Shafer triple for Ω ; (ii) G is weakly measurable on $\widehat{S} \times \Theta$; and (iii) $\theta \mapsto m_\theta$ is measurable. The theory is *incomplete* when $\nu_\theta^\infty(\cdot)$ defined as in (1.5) is nonadditive for a μ -nonnegligible set of parameters θ , which is the case when $G(\cdot | \theta)$ is multi-valued sufficiently often. Each theory induces a belief function κ defined as in (1.5)–(1.6). (Lemma C.1 provides

the technical measurability details needed to show that κ is well-defined.) The theory *represents* the preference \succeq if the latter has a Choquet expected utility function with belief function κ .

A specific theory for the entry game in Section 1.2 should be apparent from the description given there; in particular, the parameter space is $\Theta = [0, 1]$, where η , the unknown parameter appearing in the payoff matrix, is also the unknown parameter in the theory, and the correspondence G is defined by (1.11). The theory is incomplete because it leaves the selection mechanism unspecified.

The theory representing a given preference is not unique. However, there is a canonical theory that is uniquely determined by preference. Refer to a theory as *canonical* if: $\widehat{S} = \mathcal{K}(S)$, the set of compact subsets of S (endowed with the Hausdorff metric), $\Theta = Bel(S)$, and if $G(\cdot | \theta)$ is given by

$$G(K | \theta) = K \text{ for every } K \in \mathcal{K}(S) \text{ and } \theta \in \Theta.$$

Thus canonical theories differ only in the specifications of $m = (m_\theta : \theta \in \Theta)$ and μ . But two canonical theories represent the same preference if and only if they share the identical m and μ .¹¹

Our axioms characterize the set of preferences that can be represented by some theory, or equivalently, by a canonical theory.

Corollary 3.2. *The following statements are equivalent:*

- (a) \succeq satisfies Belief Function Utility, Exchangeability and WOI.
- (b) \succeq can be represented by a canonical theory.
- (c) \succeq can be represented by a theory.

3.3. Updating

Because a parameter is common to all experiments, it is an element that the individual can hope to learn about by observing the outcomes of some experiments. In the exchangeable Bayesian model, inference or learning are modeled by Bayesian updating of the prior μ over the parameter space $\Delta(S)$; and its appealing justification, which consists of the requirements of consequentialism (conditional preference does not depend on contingencies that were possible ex ante but that were not realized) and dynamic consistency, adds to the normative appeal of the model. Our model also admits an extension to include updating of the prior μ on the relevant parameter space Θ , the one appearing in the theory representing the individual's preference. The appropriate model of updating is the one described in Epstein and Seo [15], whose main features are outlined briefly here for the convenience of the reader. We restrict attention to canonical theories, where the parameter space is $Bel(S)$, though it applies to all other theories as well.

First we explain the connection between the present model and that in our earlier paper. The (ex ante) utility function in the latter has the form

$$U(f) = \int_{\mathcal{V}} V(f) d\mu'(V), \tag{3.6}$$

¹¹ Let (m, μ) and (m', μ') represent the same preference \succeq . Then, \succeq satisfies the axioms in Theorem 3.1(a) and hence $\mu = \mu'$. Further, with μ -probability 1, m_θ and m'_θ determine the same belief function via (A.1), and thus $m_\theta = m'_\theta$ μ -a.s. by Theorem A.1 (the Choquet Theorem).

where \mathcal{V} , a subset of all maxmin utility functions on acts over Ω , is the class of so-called “IID utility functions.” For the present model, it follows from [Theorem 3.1](#), that utility has the form

$$U(f) = \int_{Bel(S)} V_v(f) d\mu(v), \quad (3.7)$$

where V_v , defined in [\(3.4\)](#), is both in the Gilboa–Schmeidler class and (as can be verified) is an IID utility function. Thus [\(3.7\)](#) is the special case of [\(3.6\)](#) where μ' has support in the (strict) subset of all IID utility functions that also conform to belief function utility. Because our model is a special case, the updating rule derived in (2010) can be applied also here.

Assume that choices are made not only ex ante, but also after observing the outcomes of the first n experiments. Assume also that conditional preference after observing the finite sample $s^n = (s_1, \dots, s_n)$ of outcomes also satisfies the axioms in [Theorem 3.1](#); thus it also has a utility function as described in the theorem though with a suitable posterior $\mu(\cdot | s^n)$. Think of the entry game for concreteness, where priors and posteriors are over the value of η . Straightforward application of Bayes' Rule is not possible because the likelihood of observing T (two firms enter) conditional on η could be any number in the interval $[0, \eta]$; or put another way, the individual's theory of the entry game setting does not imply a unique likelihood function. However, in (2010) we show that under two assumptions, the individual should apply Bayes' Rule to her prior using some likelihood function that is subjective just as is her prior. An example is the likelihood function obtained by averaging over the interval $[0, \eta]$ so that

$$L(s^n | \eta) = \int_{q=0}^{\eta} q^{\#\{i:s_i=T\}} (1-q)^{\#\{i:s_i \neq T\}} dq.$$

Further, the above prescription for updating follows from two simple axioms: consequentialism, as in the classical model, and a weakened form of dynamic consistency. The latter axiom weakens dynamic consistency because it imposes consistency of ex ante and conditional preferences only in situations where the PM observes outcomes in some markets and then ‘bets’ on outcomes in others. In other words, the outcomes in markets 1 to n are ‘pure’ signals and are not payoff relevant, while outcomes in markets $n + 1$ and beyond influence payoffs but are not a source of information for further updating (which is done only once). We emphasize also that the identical axioms characterize such an updating rule also in the present case where belief function utility, rather than maxmin, defines the framework—the identical proof applies.

Finally, we note some intuitive implications of the updating rule that reflect the role of parameters when the surrounding theory is incomplete. First, the individual *learns only about parameters* (for example, η in the entry game) and does not even attempt to learn about the idiosyncratic factors (for example, selection) affecting experiments. She does not understand the latter well enough to even theorize about them and thus does not try or expect to learn about them. Accordingly, in contrast with the Bayesian decision-maker, she does not expect to learn enough to permit probabilistic predictions about remaining experiments. Second, as illustrated in our earlier paper, in general *she may not learn the true parameter* (say η) even asymptotically in large samples. This is because the empirical distinction between parameters $\eta \neq \eta'$ is clouded by the vagaries of selection, and hence parameters are only “partially identified.”¹²

¹² See Tamer [\[39\]](#) for a survey of econometric literature on partial identification.

3.4. Frequencies

The normative value of de Finetti’s specialization of subjective expected utility is well-known: in the urns context, for example, a Bayesian agent who judges the urns to be exchangeable need only formulate a prior over $[0, 1]$, that is, over the proportion of red in the fixed part of each urn, in order to generate a predictive prior over the sequence of draws. In the absence of exchangeability and de Finetti’s representation, she would face the more daunting challenge of directly forming a predictive prior over $\{R, B\}^\infty$. In the same way, our theorem can simplify the task of an individual who is less confident in her view, or theory, of the urns and who wishes to maximize belief function utility for some belief function on $\{R, B\}^\infty$. If she accepts the axioms Exchangeability and Weak Orthogonal Independence, then she need not attempt to arrive directly at such a belief function—it suffices that she can form a prior over pairs (θ_1, θ_2) , $0 \leq \theta_1 \leq \theta_2 \leq 1$, interpreted as the minimum and maximum proportions of red in all urns, and then apply the representation in [Theorem 3.1](#).

Furthermore, though the theorem does not explicitly describe how to arrive at the prior μ , there is a way for the individual to calibrate her prior beliefs if she can manage the arguably weaker task of assessing how much she would be willing to pay for bets on empirical frequencies in large samples. Kreps [[26, Ch. 11](#)] describes the corresponding procedure for calibrating prior beliefs in the exchangeable Bayesian model and argues for its usefulness in a normative context. We see no reason to view as less useful the extended calibration method that follows.

We illustrate calibration here for the entry game example ([Section 1.2](#)) and give a general result in [Appendix D](#). Denote by $\Psi_n(\omega)$ the empirical frequency of the outcome T (two firms enter) in the first n experiments given the sample ω .

Begin with PM facing [\(1.10\)](#) who is certain that the selection probability of both firms entering the market is q in each market and i.i.d. across markets. She maximizes subjective expected utility with an exchangeable predictive prior. Therefore, the classical Law of Large Numbers (LLN) for exchangeable measures implies certainty that the empirical frequency of T converges to $q\eta$, and further that prior beliefs about η and the (certainty equivalent) utility for bets about empirical frequencies are related by¹³

$$\mu(\{\eta : 0 \leq \eta \leq b\}) = U(\{\omega : \lim \Psi_n(\omega) \leq qb\}). \tag{3.8}$$

Therefore, PM can calibrate her prior μ over the parameter η if she can arrive at certainty equivalents for the indicated bets on limiting empirical frequencies.

Now suppose that PM takes seriously her ignorance about equilibrium selection in that, given η , she views any number in $[0, \eta]$ as a possible probability for T being selected when there are multiple equilibria. Then, because she is uncertain about how selection may differ and be correlated across markets, she is not certain that empirical frequencies converge. Nevertheless, there exists the following connection between prior beliefs about η and her certainty equivalents for suitable bets on empirical frequencies¹⁴:

$$\mu(\{\eta : 0 \leq \eta \leq b\}) = U(\{\omega : \limsup \Psi_n(\omega) \leq b\}). \tag{3.9}$$

¹³ $\{\omega : \lim \Psi_n(\omega) \leq q\bar{\eta}\}$ denotes both the event and the bet on the event with winning and losing prizes 1 and 0. Similarly below.

¹⁴ The proof (see [Appendix D](#)) is based on a LLN for i.i.d. products of belief functions due to Maccheroni and Marinacci [[28](#)].

In other words, the probability assigned to values of η no greater than b equals the certainty equivalent of the bet (with prizes 1 and 0) that, for all $\delta > 0$, the empirical frequency of T is less than $b + \delta$ in all sufficiently large samples.

Only the lim sup of empirical frequencies appears above because we assumed that η is associated with the probability interval $[0, \eta]$ for the outcome T , having zero as its left endpoint. More generally, if she is not certain that the minimum probability of T is 0, then she entertains intervals of the form $[\underline{\eta}, \eta]$ as the unknown parameter, and forms a prior over the parameter space $\{(\underline{\eta}, \eta) : 0 \leq \underline{\eta} \leq \eta \leq 1\}$. Then (3.9) generalizes to:

$$\mu(\{(\underline{\eta}, \eta) : a \leq \underline{\eta} \leq \eta \leq b\}) = U(\{\omega : [\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset [a, b]\}).$$

See Appendix D for a more general result and proof.

4. Two applications

We illustrate our model and its tractability by applying it to two classical problems—prediction and portfolio choice.

4.1. Prediction

Consider an optimal point prediction problem, that of predicting optimally the empirical frequency of each outcome when the experiment has two possible outcomes, denoted R and B as in the urns example. We model optimal prediction by the following decision problem:

$$\max_{\alpha \in [0,1]} \int_{Bel(S)} \int_{\Omega} G(\Psi_n(\omega) - \alpha) d\nu^\infty d\mu(\nu), \tag{4.1}$$

where $-G$ is a bounded strictly convex loss function that penalizes large differences between the predicted and realized frequencies α and $\Psi_n(\omega)$ respectively for the outcome R .

Given a belief function ν on S , define $\nu^*(R) = 1 - \nu(B)$, the maximum probability of R . Thus ν can be identified with the probability interval $[\nu(R), \nu^*(R)]$ for R .

Theorem 4.1. *There is a unique maximizer α_n in (4.1) and $\alpha_\infty \equiv \lim_{n \rightarrow \infty} \alpha_n$ exists. Moreover,*

$$\{\alpha_\infty\} = \arg \max_{\alpha \in [0,1]} \int \min \{G(\nu(R) - \alpha), G(\nu^*(R) - \alpha)\} d\mu(\nu). \tag{4.2}$$

The limiting prediction α_∞ serves as an approximately optimal prediction for a sufficiently large number of experiments. Intuition for its characterization via (4.2) is derived from the LLN for i.i.d. belief functions (see (D.1) and (D.2)). Fix ν and α and consider

$$\int_{\Omega} G(\Psi_n(\omega) - \alpha) d\nu^\infty = \min_{P \in core(\nu^\infty)} \int_{\Omega} G(\Psi_n(\omega) - \alpha) dP. \tag{4.3}$$

The LLN implies that limit points of empirical frequencies are certain to lie in the interval $[\nu(R), \nu^*(R)]$, and that for some possible probability law they are certain to be found arbitrarily near an endpoint; that is, for any $\nu(R) < a < b < \nu^*(R)$,

$$P(\{[\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset [a, b]\}) = 0$$

for some P in $core(v^\infty)$. This suggests that, when n is large, for the worst-case scenario in (4.3) it suffices to consider only samples that have empirical frequency equal to one of $v(R)$ and $v^*(R)$, as in (4.2).

To gain further insight into the nature of optimal predictions, we specialize the model by adding three assumptions. First, let the penalty function G be quadratic,

$$G(t) = -t^2.$$

Second, suppose as in the entry game that the only relevant belief functions are of the form v_η satisfying

$$v_\eta(R) = 0 \text{ and } v_\eta^*(R) = \eta.$$

Finally, assume certainty that the true parameter value η is known. Then (4.2) yields the closed-form solution

$$\alpha_\infty = \eta/2.$$

At the other extreme of predictions for a small number of markets, elementary calculations yield:

$$\alpha_1 = \begin{cases} \eta & \text{if } \eta \leq \frac{1}{2} \\ \frac{1}{2} & \text{if } \frac{1}{2} \leq \eta \end{cases} \tag{4.4}$$

and

$$\alpha_2 = \begin{cases} \eta & \text{if } \eta \leq \frac{1}{4} \\ \frac{1}{4} & \text{if } \frac{1}{4} \leq \eta \leq \frac{1}{2} \\ \eta^2 & \text{if } \frac{1}{2} \leq \eta \leq \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \text{if } \frac{1}{\sqrt{2}} \leq \eta \end{cases} \tag{4.5}$$

One observation is that $\alpha_1 \neq \alpha_2 \neq \alpha_\infty$. Thus the optimal prediction depends on the number of experiments being considered, in contrast to a Bayesian with exchangeable predictive prior whose prediction for any number of experiments would be the probability of R . It follows that the pair of predictions for both one and two experiments cannot be replicated by the exchangeable Bayesian model.

The prediction for two experiments (urns or markets) reveals the influence of ambiguous correlation in a more explicit way. By the appropriate form of (A.5), the optimal prediction problem (when $\mu(v) = 1$) can be rewritten in the form

$$\max_{\alpha \in [0,1]} \min_{P \in core(v^\infty)} \int_{\Omega} G(\Psi_n(\omega) - \alpha) dP.$$

Then it follows from the minimax theorem that α_n is optimal if and only if it solves

$$\max_{\alpha \in [0,1]} \int_{\Omega} G(\Psi_n(\omega) - \alpha) dP^*,$$

where P^* is a worst-case scenario for α_n , that is, it solves

$$\min_{P \in core(v^\infty)} \int_{\Omega} G(\Psi_n(\omega) - \alpha_n) dP.$$

In brief, one can view α_n as the best response to the scenario P^* , and thus by identifying P^* we can understand the reasons for the choice of α_n . Apply the preceding to α_2 in (4.5). The corresponding worst-case measure P^* satisfies¹⁵:

$$\begin{aligned} mrg_{\{1,2\}}P^* &= \begin{pmatrix} \eta^2 & 0 \\ 0 & 1 - \eta^2 \end{pmatrix} \text{ if } \eta > \frac{1}{2}, \\ mrg_{\{1,2\}}P^* &= \begin{pmatrix} \eta^2 & \eta(1 - \eta) \\ \eta(1 - \eta) & (1 - \eta)^2 \end{pmatrix} \text{ if } \eta < \frac{1}{2}. \end{aligned} \tag{4.6}$$

Therefore, for η larger than $\frac{1}{2}$, the optimal prediction responds to the worst-case concern that selection is positively correlated across markets (if R is realized in one urn, then it is certain to be realized also in the other, and similarly for B). Correlation does not play a role when predicting given $\eta < \frac{1}{2}$, where P^* is an i.i.d. product on $S_1 \times S_2$. As noted above, the predictions of our model are distinct from those of the exchangeable Bayesian model when considering both one and two experiments. In addition, even if we consider only the predictions for two experiments, though a Bayesian using P^* would also predict as in (4.5), when $\eta > \frac{1}{2}$, it might seem unnatural for an analyst to have complete confidence that $P^*(R_1 B_2) = P^*(B_1 R_2) = 0$.

4.2. Portfolio diversification

Given a safe asset and n stocks with uncertain returns, when is it optimal to diversify and hold stocks? It is well-known that under subjective expected utility maximization, nondiversification is a knife-edge property. In the case of one uncertain asset, Dow and Werlang [11] show that when there is ambiguity about stock returns, then nondiversification is optimal for a range of stock prices or “expected” excess returns. A common response to this result, based on the intuition suggested by the classical LLN, is to conjecture that with a large number of stocks, diversification would diminish the effect of ambiguity, the noted range would be reduced, and the expected utility result would be restored asymptotically for large n . We show that this conjecture is false given our model of preference and a suitable specification of the set of securities. Under these conditions, the decision whether or not to hold a nonzero amount of a given stock does not depend on how many other stocks are available.

Let the safe asset have the constant return r and let the i th stock have excess return X_i , where, for $i = 1, 2, \dots, n$,

$$X_i = a_i + \sigma_i \epsilon_i,$$

with a_i and $\sigma_i > 0$ being known constants. Each residual ϵ_i is random and takes on values in a compact interval S of the real line. Beliefs about each ϵ_i are represented by a common belief function ν on S . Further, beliefs about residuals jointly are represented by the belief function ν^n defined as the following marginal of ν^∞ :¹⁶

$$\nu^n(A) \equiv \nu^\infty(A \times S^\infty), \text{ for every } A \subset S^n.$$

¹⁵ $mrg_{\{1,2\}}P^*$ denotes the marginal of P^* on $S_1 \times S_2$. In each matrix, the first row gives the probabilities of $R_1 R_2$ and $R_1 B_2$ in that order, and the probabilities of $B_1 R_2$ and $B_1 B_2$ constitute the second row. The proof uses the characterization (A.8) of $\text{core}(\nu^\infty)$.

¹⁶ ν^{n-1} , used below, is defined similarly.

A special case is where the residuals are determined by draws from a sequence of urns as in Section 1.2, though finite, and where $\epsilon_i = \text{high}$ if red is drawn from the i th urn and $= \text{low}$ otherwise. Note that the residuals are exchangeable, but that stock returns are not because of possibly differing location and scale parameters a_i and σ_i .

Denote by w_i the number of shares of stock i in the portfolio, $w = (w_1, \dots, w_n) \in \mathbb{R}^n$. Normalize total investor wealth to equal 1. Then a portfolio is chosen to maximize $U(w)$, where

$$U(w) = \int_{S^n} u(\sum_i w_i X_i + r) dv^n.$$

The utility index u is assumed to be increasing, concave and differentiable.

The Dow–Werlang result implies that when $n = 1$, the investor will not participate in the stock market if and only if

$$\int X_1 dv \leq 0 \leq - \int -X_1 dv. \tag{4.7}$$

The interpretation is that the minimum expected excess return (taken over measures in the core of ν) being nonpositive dictates against going long, while the maximum expected excess return being nonnegative dictates against going short. Here we show that under the above assumptions, the identical condition is necessary and sufficient for $w_1^* = 0$ also when there are n stocks. Thus the possibility of diversification does not promote participation. For example, when stock returns are related to draws from urns as above, then stocks do not hedge one another because there is no basis for believing that the idiosyncratic components of urns are related in any particular way. There is evidently a close connection to Weak Orthogonal Independence, though the axiom applies to Anscombe–Aumann acts (or assuming risk neutrality) and thus does not directly imply the nonparticipation result. For very large n , there is also an intuitive connection to the LLN for belief functions cited in the previous subsection, whereby empirical frequencies do not converge to a point.

The proof is straightforward. Assume (4.7). Let w^* be an optimal portfolio, $w_{-1}^* = (w_2^*, \dots, w_n^*)$, $\epsilon_{-1} = (\epsilon_2, \dots, \epsilon_n)$ and $X_{-1} = (X_2, \dots, X_n)$. Then, for any $w_1 \geq 0$, we have

$$\begin{aligned} U(w_1, w_{-1}^*) &= \int \int u(w_1 X_1 + \sum_{i>1} w_i^* X_i + r) dv(\epsilon_1) dv^{n-1}(\epsilon_{-1}) \\ &\leq \int u\left(w_1 \left(\int X_1 dv\right) + \sum_{i>1} w_i^* X_i + r\right) dv^{n-1}(\epsilon_{-1}) \\ &\leq \int u(0 + \sum_{i>1} w_i^* X_i + r) dv^{n-1}(\epsilon_{-1}) = U(0, w_{-1}^*). \end{aligned}$$

The first equality is by the Fubini property for belief functions (Ghirardato [18]), the first inequality is by a version of Jensen’s inequality (Dow and Werlang [11], for example), and the next inequality is by (4.7) and $w_1 \geq 0$. Thus, $w_1 = 0$ is optimal under the condition $w_1 \geq 0$. A similar argument shows that $w_1 = 0$ is optimal under the condition $w_1 \leq 0$. Thus $w_1^* = 0$. Proof of the converse relies on Dow and Werlang’s Lemma 4.1 and is omitted.

5. Concluding remarks and related literature

The celebrated de Finetti result is important because: (a) the representation admits an intuitive interpretation; (b) the characterizing axiom is simple and transparent; (c) the model admits a

theory of updating; and (d) it provides a behavioral foundation for modeling a decision-maker (DM) as being uncertain about an unknown “parameter” that describes common features of every experiment and trying to learn about them. Our critique is that de Finetti models a decision-maker whose theory of her environment is complete in the sense that once she learns the parameter value, then she knows the probability law describing the sequence of experiments which is all she needs to make predictions and decisions regarding future experiments. Moreover, under suitable assumptions, she is certain that she will indeed learn the true parameter asymptotically given enough data. Accordingly, the theory leaves no room for doubt about what is well-understood or for more modest ambitions about what is learnable. Our primary contribution is to generalize de Finetti so as to accommodate less extreme self-confidence (or naivete) while retaining much of (a)–(d).

Our earlier paper (2010) is closely related.¹⁷ It introduces the issue of ambiguity about heterogeneity and correlation across experiments and adopts the framework of maxmin utility (Gilboa and Schmeidler [19]) to model it. In addition, the functional form for utility characterized here is a special case of a model studied there (see (3.6) and (3.7)). Therefore, we want to make perfectly clear the value-added here. It does not lie in the modeling of updating or learning, where we simply cite the earlier work (Section 3.3). However, in other respects, the model in (2010) is limited at the levels of both functional form and axiomatics.¹⁸

Here is an outline of the value-added over (2010) in each of the dimensions (a), (b) and (d). (a) The representation in (2010), given in (3.6) above, is not standard in that it describes preferences satisfying some axioms as a mixture of “IID utility functions,” but little is known about the class \mathcal{V} of IID utility functions beyond their definition and two examples (one of which is the model studied here). Therefore, it is not clear what the representation tells us about the given preference order. Here the representation is as concrete and explicit as de Finetti’s. (b) The primary axioms in (2010) are Orthogonal Independence and Super-Convexity, neither of which is “transparent”; see the latter in particular and compare both with our main axiom here WOI. (d) The improvements herein noted in (a) and (b) have implications with regard to the behavioral foundations provided for parameters. The earlier paper interprets various representations in terms of uncertainty about unknown parameters. In the context of a special case of the general model, (see (5.5), p. 332), the unknown parameter takes the concrete and natural form of a set of likelihoods, probability measures on S , much as in this paper. However, the axiomatic foundations for that special model are unknown. The general axiomatically-based representation in (2010), given above in (3.6), is also described there in terms of uncertainty about a parameter, the unknown IID utility function V . However, as noted above, the axioms are much less transparent and little is known about these functions V , which limits their appeal and usefulness as parameters. (For instance, in the urns example, it is not clear what it would mean to think of IID utility functions as parameters.) Only the present paper provides *both* behavioral foundations *and* a concrete natural notion of parameter. Our individual perceives common elements—captured by belief functions on the outcome set of a single experiment, or by probability intervals when each experiment is binary—but who recognizes that the parameters alone do not capture all that is going on, for example, there is also poorly understood equilibrium selection in the entry game. Policy choice in the context of an entry game with multiple equilibria is one example where a belief function

¹⁷ There are two models in (2010). We refer to the more relevant one summarized in Thm. 5.2.

¹⁸ The following comparison with the 2010 paper is included in order to respond to an editor’s questions about the value-added herein.

is a natural choice for modeling a decision-maker with such an incomplete theory, but there are many others. We view this as a main contribution of the paper.¹⁹

Shafer [37] is the first, to our knowledge, to discuss the use of belief functions within the framework of parametric 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 μ models beliefs about ℓ , the unknown 'parameter'. An obvious generalization is to replace μ by a belief function on $\Delta(S)$, or more generally by a set of probability measures on $\Delta(S)$, thus generalizing prior beliefs. Such a model is axiomatized within the maxmin framework in our (2010) paper, (where it is the first of the two models studied), and other models in this spirit are studied by Al-Najjar and de Castro [2], Cerreia-Vioglio et al. [6] and Klibanoff, Mukerji and Seo [25]. In all cases, the functional form for utility can be interpreted in terms of an unknown parameter—the ambiguous probability law for each experiment. A difference from our model is that the decision-maker knows everything she needs once she knows the parameter—her theory is complete—just as in de Finetti's model.²⁰ At a behavioral level, these models are distinguished from ours because they cannot accommodate either of the rankings (2.3) and (2.4), which we take as canonical illustrations of aversion to ambiguity about heterogeneity and correlation respectively.

Epstein and Schneider [13,14] study similar issues to those that concern us here in a setting where experiments are ordered in time. One difference is that their analysis is not axiomatic—they suggest functional forms and provide informal justification primarily through applications. Another difference is that their models are recursive and because of that they violate Exchangeability (thus excluding the functional form studied here), which violation is at least plausible in a temporal setting where experiments are distinguished by the time at which they are run, but in our view, makes them inappropriate for a cross-sectional setting which is our focus here.

Finally, consider connections to the applied literature on entry games and the broader literature on partial identification. Most of the latter literature studies inference and estimation. Our presumption, however, is that, as stated by Tamer [39, p. 174]: "One main motivation for empirical work in economics is to evaluate policies, with an important purpose of decision making." The only papers of which we are aware that explicitly address policy choice in the context of partially identified models are Manski [29–31] and Kasy [24]. Their approaches are not axiomatic and their models are much different than ours, in particular, they do not follow in the footsteps of de Finetti.

Appendix A. Belief functions

The following notation is used throughout the appendices. For any compact metric space Ω , $\mathcal{K}(\Omega)$ is the space of compact subsets endowed with the Hausdorff metric; $\Delta(\Omega)$ is the space of Borel countably additive probability measures on Ω endowed with the weak convergence

¹⁹ Another "first," though one that is more narrowly of interest within decision theory, is that we are the first to provide behavioral foundations for a particular i.i.d. product rule for belief functions (see Remark 1).

²⁰ Marinacci [32] shows that, within the multiple-priors framework and under suitable assumptions about updating, the decision-maker is certain ex ante that she will learn the true law asymptotically if she draws one ball from each of many urns. The intuition is that given her perception or theory of the urns, it is as though she were sampling with replacement from a single urn.

topology; and $Bel(\Omega)$ is the space of belief functions endowed with the topology for which $\kappa_n \rightarrow \kappa$ if and only if $\int f d\kappa_n \rightarrow \int f d\kappa$ for every continuous function f on Ω , where the integral is in the sense of Choquet. All three spaces are compact metric. They are endowed with the corresponding Borel σ -algebras. For any metric space X , its σ -algebra is denoted Σ_X .

This appendix collects some facts about belief functions that support assertions in the text and in the proofs below. We deal with belief functions on Ω , which until further notice can be any compact metric space.

A belief function is most commonly defined as a set function $\kappa : \Sigma_\Omega \rightarrow [0, 1]$ satisfying:

- Bel. 1 $\kappa(\emptyset) = 0$ and $\kappa(\Omega) = 1$
- Bel. 2 $\kappa(A) \leq \kappa(B)$ for all Borel sets $A \subset B$
- Bel. 3 $\kappa(B_n) \downarrow \kappa(B)$ for all sequences of Borel sets $B_n \downarrow B$
- Bel. 4 $\kappa(G) = \sup\{\kappa(K) : K \subset G, K \text{ compact}\}$, for all open G
- Bel. 5 κ is totally monotone (or ∞ -monotone): for all Borel sets B_1, \dots, B_n ,

$$\kappa\left(\bigcup_{j=1}^n B_j\right) \geq \sum_{\emptyset \neq J \subset \{1, \dots, n\}} (-1)^{|J|+1} \kappa\left(\bigcap_{j \in J} B_j\right).$$

These conditions are adapted from Philippe et al. [34]. Conditions Bel. 1–Bel. 4 form a common definition of capacity (Schmeidler [35]). When restricted to probability measures, Bel. 4 is the well-known property of regularity. If the inequalities in Bel. 5 are restricted to $n = 2$, one obtains that κ is convex (supermodular, or 2-alternating).

An important result regarding belief functions is the Choquet Theorem. Our statement of the theorem relies on Philippe et al. [34, Thms. 2 and 3], Molchanov [33, Thm. 5.1] and Castaldo et al. [5, Thm. 3.2]. Note that, by Philippe et al. [34, Lemma 1], $\{K \in \mathcal{K}(\Omega) : K \subset A\}$ is universally measurable for every $A \in \Sigma_\Omega$. Further, any Borel probability measure (such as m on Borel subsets of $\mathcal{K}(\Omega)$) admits a unique extension (also denoted m) to the collection of all universally measurable sets.²¹

Theorem A.1 (Choquet). *The set function $\kappa : \Sigma_\Omega \rightarrow [0, 1]$ satisfies Bel. 1–Bel. 5 if and only if there exists a (necessarily unique) Borel probability measure M on $\mathcal{K}(\Omega)$ such that*

$$\kappa(A) = M(\{K \in \mathcal{K}(\Omega) : K \subset A\}), \text{ for every } A \in \Sigma_\Omega. \tag{A.1}$$

Moreover, in that case, for every measurable $f : \Omega \rightarrow [0, 1]$, the Choquet integral $\int_\Omega f d\kappa$ satisfies:

$$\begin{aligned} \int_\Omega f d\kappa &= \int_{\mathcal{K}(\Omega)} \left(\inf_{P \in \Delta(K)} P \cdot f \right) dM(K) \\ &= \int_{\mathcal{K}(\Omega)} \left(\inf_{x \in K} f(x) \right) dM(K). \end{aligned} \tag{A.2}$$

²¹ Throughout, given any Borel probability measure, we identify it with its unique extension to the σ -algebra of universally measurable sets. Below $P \cdot f$ is short-hand for $\int_X f dP$.

We use frequently below the implication that every belief function (as defined by Bel. 1–Bel. 5) on a space Ω can be identified with a unique probability measure on the space of its closed subsets; in fact, $Bel(\Omega)$ is homeomorphic to $\Delta(\mathcal{K}(\Omega))$. Another implication is that the definition via Bel. 1–Bel. 5 is equivalent to that given in the text via (1.3) and due to Dempster [8] and Shafer [36]. For one direction, the theorem proves that Bel. 1–Bel. 5 imply the representation (A.1), which is the special case of (1.3) where $\widehat{\Omega} = \mathcal{K}(\Omega)$ and Γ maps any K (a point in $\mathcal{K}(\Omega)$) into K (a subset of Ω). Conversely, let κ be defined via the triple $(\widehat{\Omega}, M, \Gamma)$ and (1.3). View Γ as a function from $\widehat{\Omega}$ into $\mathcal{K}(\Omega)$. Then Γ is measurable (Aliprantis and Border [1, Thm. 18.10]) and induces the measure $M' = M \circ \Gamma^{-1}$ on $\mathcal{K}(\Omega)$. Then the Choquet Theorem implies that $\kappa(\cdot) = M \circ \Gamma^{-1}(\{K : K \subset \cdot\})$ satisfies Bel. 1–Bel. 5 and $M' = M$.

Associated with any belief function κ is its core defined by

$$core(\kappa) = \{P \in \Delta(\Omega) : P(\cdot) \geq \kappa(\cdot)\}.$$

Then²²

$$core(\kappa) = \left\{ P \in \Delta(\Omega) : P = \int_{\widehat{\Omega}} p_{\widehat{\omega}} dM(\widehat{\omega}), p_{\widehat{\omega}} \in \Delta(\Gamma(\widehat{\omega})) \text{ } M\text{-a.e.} \right\}. \tag{A.3}$$

Turn to the corresponding utility function. The objects of choice are (Borel measurable) acts $f : \Omega \rightarrow [0, 1]$, which are restricted to have finite range. The utility $U(f)$ of any act f is defined by (2.1). By Molchanov [33, Thm. 5.1], it can be expressed alternatively in the form

$$U(f) = \int_{\widehat{\Omega}} \left(\inf_{\omega \in \Gamma(\widehat{\omega})} f(\omega) \right) dM(\widehat{\omega}). \tag{A.4}$$

This expression for utility reflects the individual’s perception that given the auxiliary state $\widehat{\omega}$, the true payoff relevant state lies in $\Gamma(\widehat{\omega})$ but there is ignorance within $\Gamma(\widehat{\omega})$. Put another way, the marginal distribution of the subsets $\{\Gamma(\widehat{\omega})\}$ is given by M , but conditional distributions within each $\Gamma(\widehat{\omega})$ are unrestricted.

Belief function utility is a special case of the maxmin model (Gilboa and Schmeidler [19]) with set of priors equal to $core(\kappa)$:

$$U(f) = \min_{P \in core(\kappa)} \int_{\Omega} f dP. \tag{A.5}$$

Accordingly, it inherits the following properties that play a central role in the multiple-priors model: For all acts f and g , and for all constants x ,

$$U(\alpha x + (1 - \alpha)g) = \alpha x + (1 - \alpha)U(g), \tag{A.6}$$

and

$$U(\alpha f + (1 - \alpha)g) \geq \alpha U(f) + (1 - \alpha)U(g). \tag{A.7}$$

Gilboa and Schmeidler [19] refer to these properties as certainty additivity and ambiguity aversion respectively. We use them repeatedly.

²² When the support of M is not finite, a measurability assumption for $\widehat{\omega} \mapsto p_{\widehat{\omega}}$ must be added to give meaning to this expression.

As noted, the preceding applies to any state space. Now we consider further structure that is relevant in a setting with repeated experiments. Thus consider a sequence of experiments indexed by the set \mathbb{N} of positive integers. Each experiment yields an outcome in S (a compact metric space). Uncertainty concerns the outcomes of all experiments, and thus let Ω be defined by

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

Let $\nu \in \text{Bel}(S)$ be generated by (\widehat{S}, m, G) . We defined ν^∞ to be the belief function on Ω represented by $(\widehat{S}^\infty, m^\infty, G^\infty)$, where m^∞ is the ordinary i.i.d. product of the probability measure m , and G^∞ is the correspondence $G^\infty : \widehat{S}^\infty \rightsquigarrow \Omega = S^\infty$ given by (3.1). The Choquet Theorem gives an alternative characterization of the product that we use frequently. In particular, it implies that the product ν^∞ does not depend on the particular representation (\widehat{S}, m, G) for ν .

Lemma A.2. *Let $\nu \in \text{Bel}(S)$ correspond to $m \in \Delta(\mathcal{K}(S))$ as in the Choquet Theorem. Then $\nu^\infty \in \text{Bel}(\Omega)$ is the unique belief function corresponding to $M = m^\infty \in \Delta(\mathcal{K}(\Omega))$ as in the Choquet Theorem (where m^∞ is the i.i.d. product of the measure m).*

The proof of the lemma is omitted. Note that m^∞ is a measure on $[\mathcal{K}(S)]^\infty$ which is a subset of $\mathcal{K}(\Omega)$. Therefore, it can be identified with a measure on $\mathcal{K}(\Omega)$.

The core of ν^∞ has the following characterization: By Philippe et al. [34, Thm. 3],

$$\text{core}(\nu^\infty) = \int \Delta(G(\widehat{s}_1) \times G(\widehat{s}_2) \times \dots) dm^\infty(\widehat{s}_1, \dots) \tag{A.8}$$

where the integral is an Aumann integral. As noted following Theorem 3.1, the presence of the entire simplices $\Delta(G(\widehat{s}_1) \times G(\widehat{s}_2) \times \dots)$ implies that the core contains both nonidentical products and nonproduct measures.

Some of the preceding is illustrated by the following verification of (3.5). Specifically, we show that

$$\begin{aligned} V_\nu\left(\frac{1}{2}f + \frac{1}{2}\pi f\right) &\equiv \int \left(\frac{1}{2}f + \frac{1}{2}\pi f\right) d\nu^\infty \\ &= (v(R))^2 + (v(B))^2 + v(R)v(B)(1 - v(R) - v(B)), \end{aligned}$$

where f denotes the bet $\{R_1R_2, B_1B_2\}$ and π permutes the second and third markets. Let ν on $S = \{R, B\}$ be given by

$$v(R) = \theta_1, \quad v(B) = 1 - \theta_2, \quad \theta_1 < \theta_2.$$

Adopt the Dempster–Shafer representation as in (1.7)–(1.9) except that, for greater notational clarity, let $\widehat{S} = \{r, b, u\}$, $m(r) = \theta_1$, $m(b) = 1 - \theta_2$, and define G in the obvious way. By the definition of Choquet integration,

$$\int \left(\frac{1}{2}f + \frac{1}{2}\pi f\right) d\nu^\infty = v^\infty(E_1) + \frac{1}{2}[v^\infty(E_1 \cup E_2) - v^\infty(E_1)],$$

where

$$\frac{1}{2}f + \frac{1}{2}\pi f = \begin{cases} 1 & \text{on } E_1 = \{R_1R_2R_3, B_1B_2B_3\} \\ \frac{1}{2} & \text{on } E_2 = \{R_1R_2B_3, R_1B_2R_3, B_1B_2R_3, B_1R_2B_3\} \\ 0 & \text{on } E_3 = \{R_1B_2B_3, B_1R_2R_3\}. \end{cases}$$

By definition of ν^∞ :

$$v^\infty(E_1) = m^\infty(\{r_1 r_2 r_3, b_1 b_2 b_3\}) = (\theta_1)^3 + (1 - \theta_2)^3,$$

and

$$\begin{aligned} v^\infty(E_1 \cup E_2) &= m^\infty(\{r_1 r_2 u_3, r_1 u_2 r_3, b_1 b_2 u_3, b_1 u_2 b_3, u_1 r_2 b_3, u_1 b_2 r_3\}) \\ &= 2(\theta_1)^2 - (\theta_1)^3 + 2(1 - \theta_2)^2 - (1 - \theta_2)^3 + 2\theta_1(1 - \theta_2)(\theta_2 - \theta_1). \end{aligned}$$

Thus

$$\begin{aligned} &\int \left(\frac{1}{2}f + \frac{1}{2}\pi f\right) dv^\infty \\ &= (\theta_1)^3 + (1 - \theta_2)^3 \\ &\quad + \frac{1}{2} \left[2(\theta_1)^2 - (\theta_1)^3 + 2(1 - \theta_2)^2 - (1 - \theta_2)^3 + 2\theta_1(1 - \theta_2)(\theta_2 - \theta_1) \right] \\ &= (\theta_1)^2 + (1 - \theta_2)^2 + \theta_1(1 - \theta_2)(\theta_2 - \theta_1). \end{aligned}$$

Appendix B. Proof of Theorem 3.1

For any subset I of $\{1, 2, \dots\}$, let Σ_I denote the product σ -algebra on $\prod_{i \in I} S_i$, identified with a σ -algebra on $\Omega = S_1 \times S_2 \times \dots$. Denote by \mathcal{F}_I the set of all Σ_I -measurable acts.

Throughout the appendix, if $v \in \text{Bel}(S)$ and $\kappa \in \text{Bel}(\Omega)$, then $m_v \in \Delta(\mathcal{K}(S))$ and $M_\kappa \in \Delta(\mathcal{K}(\Omega))$, respectively, are the corresponding measures provided by the Choquet Theorem.

We use (A.2) repeatedly without reference.

First we prove the measurability required to show that the integrals in (3.3) and (3.2) are well-defined. (Recall that any Borel probability measure μ has a unique extension to the class of all universally measurable subsets.)

Lemma B.1. *The mapping $v \mapsto v^\infty(A)$ is universally measurable for any $A \in \Sigma_\Omega$.*

Proof. Since $\text{Bel}(S)$ and $\Delta(\mathcal{K}(S))$ are homeomorphic, and in light of (A.1), it is enough to prove analytical (and hence universal) measurability of the mapping from $\Delta(\mathcal{K}(S))$ to \mathbb{R} given by

$$m \mapsto \int_{[\mathcal{K}(S)]^\infty} m(\{K \in [\mathcal{K}(S)]^\infty : K \subset A\}) dm^\infty(K).$$

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

Step 2. $P \mapsto P(C)$ from $\Delta([\mathcal{K}(S)]^\infty)$ to \mathbb{R} is analytically measurable for any measurable subset C of $[\mathcal{K}(S)]^\infty$, by Bertsekas and Shreve [4, p. 169].

Step 3. $\{K \in [\mathcal{K}(S)]^\infty : K \subset A\}$ is coanalytic by Philippe et al. [34, p. 772], and hence analytically measurable.

Steps 1, 2 and 3 complete the proof. \square

Necessity of the axioms: Belief Function Utility is obvious. Note that $U(f)$ can be written as $\int V_v(f) d\mu(v)$ where V_v , defined in (3.4). We verify that V_v satisfies Exchangeability and WOI, which implies the same for U . By Lemma A.2, $M_{v^\infty} = (m_v)^\infty$ is an i.i.d. measure on $[\mathcal{K}(S)]^\infty$, hence exchangeable. Therefore,

$$\begin{aligned}
 V_v(\pi f) &= \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} \pi f(\omega) dm_v^\infty(K) = \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} f(\pi\omega) dm_v^\infty(K) \\
 &= \int_{\mathcal{K}(\Omega)} \inf_{\pi\omega \in \pi K} f(\pi\omega) dm_v^\infty(K) = \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} f(\omega) d(\pi m_v^\infty)(K) \\
 &= \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} f(\omega) dm_v^\infty(K) = V_v(f).
 \end{aligned}$$

The functional form satisfies a strengthening of WOI that is not restricted to binary acts or to $\frac{1}{2}/\frac{1}{2}$ mixtures. For simplicity, let $f \in \mathcal{F}_1, g \in \mathcal{F}_2$ and $0 < \alpha \leq 1$. Then

$$\begin{aligned}
 &V_v(\alpha f + (1 - \alpha)g) \\
 &= \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} [\alpha f(\omega) + (1 - \alpha)g(\omega)] dm_v^\infty(K) \\
 &= \int_{[\mathcal{K}(S)]^\infty} \inf_{s_1 \in K_1, s_2 \in K_2} [\alpha f(s_1) + (1 - \alpha)g(s_2)] dm_v^\infty(K_1, K_2, \dots) \\
 &= \int_{[\mathcal{K}(S)]^\infty} \alpha \left[\inf_{s_1 \in K_1} f(s_1) \right] + (1 - \alpha) \left[\inf_{s_2 \in K_2} (1 - \alpha)g(s_2) \right] dm_v^\infty(K_1, K_2, \dots) \\
 &= \alpha \int_{[\mathcal{K}(S)]^\infty} \left[\inf_{s_1 \in K_1} f(s_1) \right] dm_v^\infty(K_1, K_2, \dots) \\
 &\quad + (1 - \alpha) \int_{[\mathcal{K}(S)]^\infty} \left[\inf_{s_2 \in K_2} g(s_2) \right] dm_v^\infty(K_1, K_2, \dots) \\
 &= \alpha V_v(f) + (1 - \alpha) V_v(g).
 \end{aligned}$$

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

It follows from a similar argument that the preference satisfies the following related conditions. Say that the finitely-based acts f and g are orthogonal, written $f \perp g$, if there exist disjoint subsets $I, J \subset \mathbb{N}$ such that $f \in \mathcal{F}_I$ and $g \in \mathcal{F}_J$.

WOI*: For all $0 < \alpha \leq 1$, and all finitely-based acts f', f and g such that $f' \perp g$ and $f \perp g$,

$$f' \succeq f \iff \alpha f' + (1 - \alpha)g \succeq \alpha f + (1 - \alpha)g, \tag{B.1}$$

and its utility function U satisfies

$$U(\alpha f + (1 - \alpha)g) = \alpha U(f) + (1 - \alpha)U(g). \tag{B.2}$$

We use this characterization of WOI frequently in the sequel.

The next lemma is used in the proof of sufficiency. It shows that the axioms Exchangeability and WOI extend to apply also to nonbinary acts.

Lemma B.2. Suppose κ is a capacity on Ω and that the preference \succeq on \mathcal{F} is represented by the Choquet expected utility function U , $U(f) = \int f d\kappa$. Then:

(a) \succeq satisfies Exchangeability if and only if $f \sim \pi f$ for all finitely-based acts f and permutations π .

(b) The following statements are equivalent:

- (i) \succeq satisfies WOI*.
- (ii) U satisfies (B.2).
- (iii) \succeq satisfies WOI.
- (iv) $\nu(A \cup B) + \nu(A \cap B) = \nu(A) + \nu(B)$ for all finitely-based orthogonal events A and B .

Proof. (a) Only one direction requires proof. Assume Exchangeability, so that $\kappa(A) = \kappa(\pi A)$ for any finitely-based event A . Take a finitely-based act f . Then,

$$\begin{aligned} U(\pi f) &= \int \pi f d\kappa = \int_0^1 \kappa(\{\omega : \pi f(\omega) \geq t\}) dt \\ &= \int_0^1 \kappa(\{\omega : f(\pi\omega) \geq t\}) dt \\ &= \int_0^1 \kappa(\{\pi^{-1}(\pi\omega) : f(\pi\omega) \geq t\}) dt \\ &= \int_0^1 \kappa(\{\pi^{-1}\{\omega' : f(\omega') \geq t\}\}) dt \\ &= \int_0^1 \kappa(\{\omega' : f(\omega') \geq t\}) dt = U(f), \end{aligned}$$

where the second last equality holds because $\{\omega' : f(\omega') \geq t\}$ is finitely-based.

(b) That (ii) \implies (i) \implies (iii) is clear.

(i) implies (ii): Let f' be constant at level $U(f)$, so that $f' \sim f$. Because constant acts are orthogonal to every act, deduce that $\alpha U(f) + (1 - \alpha)g \sim \alpha f + (1 - \alpha)g$. Therefore, $\alpha U(f) + (1 - \alpha)U(g) = U(\alpha U(f) + (1 - \alpha)g) = U(\alpha f + (1 - \alpha)g)$; the first equality is due to (A.6).

(iii) implies (iv): Adapt the preceding argument to show that (iii) implies

$$U\left(\frac{1}{2}f + \frac{1}{2}g\right) = \frac{1}{2}U(f) + \frac{1}{2}U(g)$$

for all orthogonal indicator acts $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$. Then, because $\mathbf{1}_{A \cup B}$ and $\mathbf{1}_{A \cap B}$ are comonotonic,

$$\begin{aligned} \kappa(A \cup B) + \kappa(A \cap B) &= \int \mathbf{1}_{A \cup B} d\kappa + \int \mathbf{1}_{A \cap B} d\kappa = \int (\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}) d\kappa \\ &= \int (\mathbf{1}_A + \mathbf{1}_B) d\kappa = 2U\left(\frac{1}{2}f + \frac{1}{2}g\right) = U(f) + U(g) \\ &= \kappa(A) + \kappa(B). \end{aligned}$$

(iv) implies (ii): Without loss of generality, let $a > b > 0$, and write

$$a\mathbf{1}_A + b\mathbf{1}_B = b(\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}) + (a - b)\mathbf{1}_A.$$

Because $\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}$ and $\mathbf{1}_A$ are comonotonic,

$$\begin{aligned} U(a\mathbf{1}_A + b\mathbf{1}_B) &= \int b(\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}) d\kappa + \int (a - b)\mathbf{1}_A d\kappa \\ &= b(\kappa(A \cup B) + \kappa(A \cap B)) + (a - b)\kappa(A) \\ &= b(\kappa(A) + \kappa(B)) + (a - b)\kappa(A) \\ &= a\kappa(A) + b\kappa(B) = aU(\mathbf{1}_A) + bU(\mathbf{1}_B). \end{aligned} \tag{B.3}$$

This shows that $U(\alpha f + (1 - \alpha)g) = \alpha U(f) + (1 - \alpha)U(g)$ for orthogonal indicator acts f and g . To extend to any (finitely-based) orthogonal f and g , note that

$$f = a_1\mathbf{1}_{A_1} + a_2\mathbf{1}_{A_2} + \dots + a_n\mathbf{1}_{A_n}$$

for some $A_1 \subset \dots \subset A_n$ and $a_1 < \dots < a_n$. Because the acts $(\alpha a_i\mathbf{1}_{A_i} + \frac{1-\alpha}{n}g)$, $i = 1, \dots, n$, are comonotonic,

$$U(\alpha f + (1 - \alpha)g) = U\left(\sum_{i=1}^n \left(\alpha a_i\mathbf{1}_{A_i} + \frac{1-\alpha}{n}g\right)\right) = \sum_{i=1}^n U\left(\alpha a_i\mathbf{1}_{A_i} + \frac{1-\alpha}{n}g\right).$$

Similarly,

$$g = b_1\mathbf{1}_{B_1} + b_2\mathbf{1}_{B_2} + \dots + b_m\mathbf{1}_{B_m}$$

for some $B_1 \subset \dots \subset B_m$ and $b_1 < \dots < b_m$. Argue as above to derive

$$U(\alpha f + (1 - \alpha)g) = \sum_i^n \sum_j^m U\left(\frac{\alpha}{m}a_i\mathbf{1}_{A_i} + \frac{1-\alpha}{n}b_j\mathbf{1}_{B_j}\right).$$

By a slight extension of (B.3),

$$U\left(\frac{\alpha}{m}a_i\mathbf{1}_{A_i} + \frac{1-\alpha}{n}b_j\mathbf{1}_{B_j}\right) = \frac{\alpha}{m}U(a_i\mathbf{1}_{A_i}) + \frac{1-\alpha}{n}U(b_j\mathbf{1}_{B_j}).$$

Therefore,

$$\begin{aligned} U(\alpha f + (1 - \alpha)g) &= \sum_i^n \sum_j^m \left[\frac{\alpha}{m}U(a_i\mathbf{1}_{A_i}) + \frac{1-\alpha}{n}U(b_j\mathbf{1}_{B_j}) \right] \\ &= \alpha U\left(\sum_i^n a_i\mathbf{1}_{A_i}\right) + (1 - \alpha)U\left(\sum_j^m b_j\mathbf{1}_{B_j}\right) \\ &= \alpha U(f) + (1 - \alpha)U(g). \quad \square \end{aligned}$$

Sufficiency of the axioms. We apply Lemma B.2 repeatedly below without explicit reference. For example, when we invoke Exchangeability or WOI, we apply their extensions to nonbinary acts.

For $C \subset \mathcal{K}(\Omega)$, let $\pi C = \{\pi K \in \mathcal{K}(\Omega) : K \in C\}$, and for $M \in \Delta(\mathcal{K}(\Omega))$, define $\pi M \in \Delta(\mathcal{K}(\Omega))$ by $\pi M(C) = M(\pi C)$ for each Borel measurable $C \subset \mathcal{K}(\Omega)$.

Lemma B.3. For any $M \in \Delta(\mathcal{K}(\Omega))$, $M = \pi M$ for all π if and only if $M = M_\kappa$ for some exchangeable belief function κ on Ω .

Proof. If $M = M_\kappa$, then $\kappa(K) = M(\{K' \in \mathcal{K}(\Omega) : K' \subset K\})$, and

$$\begin{aligned} \kappa(\pi K) &= M(\{K' \in \mathcal{K}(\Omega) : K' \subset \pi K\}) = M(\{\pi K' \in \mathcal{K}(\Omega) : \pi K' \subset \pi K\}) \\ &= M(\{\pi K' \in \mathcal{K}(\Omega) : K' \subset K\}) = \pi M(\{K' \in \mathcal{K}(\Omega) : K' \subset K\}). \end{aligned}$$

The asserted equivalence follows because the class $\{K' \in \mathcal{K}(\Omega) : K' \subset K\}_{K \in \mathcal{K}(\Omega)}$ generates the Borel σ -algebra on $\mathcal{K}(\Omega)$. \square

Lemma B.4. Let $\kappa \in \text{Bel}(\Omega)$ and let U be the corresponding Choquet expected utility function. If U satisfies WOI, then $M_\kappa[(\mathcal{K}(S))^\infty] = 1$.

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

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

$$\exists \omega^* \in K, \text{ such that } \omega_I^* = \omega_I^1 \text{ and } \omega_J^* = \omega_J^2. \tag{B.4}$$

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 ,

$$\min_{\omega \in K} \left[\frac{1}{2}f(\omega) + \frac{1}{2}g(\omega) \right] = \frac{1}{2} \min_{\omega \in K} f(\omega) + \frac{1}{2} \min_{\omega \in K} g(\omega), \tag{B.5}$$

a.s.- M_κ : This is where WOI enters – by Lemma B.2(ii), it implies that

$$U\left(\frac{1}{2}f + \frac{1}{2}g\right) = \frac{1}{2}U(f) + \frac{1}{2}U(g).$$

Since $U(f) = \int_{\mathcal{K}(\Omega)} \inf_{\omega \in K} f(\omega) dM_\kappa(K)$, then $\int_{\mathcal{K}(\Omega)} \min_{\omega \in K} \left[\frac{1}{2}f(\omega) + \frac{1}{2}g(\omega) \right] dM_\kappa(K) = \frac{1}{2} \int_{\mathcal{K}(\Omega)} \min_{\omega \in K} f(\omega) dM_\kappa(K) + \frac{1}{2} \int_{\mathcal{K}(\Omega)} \min_{\omega \in K} g(\omega) dM_\kappa(K)$. The assertion follows from

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

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 $K \in \mathcal{K}(\Omega)$ satisfying (B.5), given f and g . Step 1 implies $m(\mathcal{B}_{f,g}) = 1$ for each $(f, g) \in \mathcal{G}$.

Step 2. $M_\kappa\left(\bigcap_{(f,g) \in \mathcal{G}} \mathcal{B}_{f,g}\right) = 1$: Since the set of continuous finitely-based acts is separable under the sup-norm topology (Aliprantis and Border [1, 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,

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

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 without loss of generality that $(f_i, g_i) \rightarrow (f, g)$. Then, by the Maximum Theorem (Aliprantis and Border [1, Thm. 17.31]),

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

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 $\{1, \dots, n\} = I \cup J$, with I and J disjoint. For each i , take closed sets

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

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 $f_i(\hat{\omega}^i) = g_i(\hat{\omega}^i) = 0$. By the construction of f_i and g_i , we have $\hat{\omega}^i \notin A_i, B_i$, which implies

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

Since $\{\hat{\omega}^i\} \subset K$ and K is compact, there is a limit point $\omega^* \in K$ satisfying (B.4).

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

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 $\omega^1, \omega^2, \dots \in K$. It suffices to show that

$$\omega^* = (\omega_1^1, \omega_2^2, \dots, \omega_n^n, \dots) \in K. \tag{B.6}$$

Since $K \in \mathcal{A}$ and $\omega^1, \omega^2 \in K$, there exists $\hat{\omega}^2 \in K$ such that $(\hat{\omega}_1^2, \hat{\omega}_2^2) = (\omega_1^1, \omega_2^2)$. Similarly, since $\hat{\omega}^2, \omega^3 \in K$, there exists $\hat{\omega}^3 \in K$ such that $(\hat{\omega}_1^3, \hat{\omega}_2^3, \hat{\omega}_3^3) = (\hat{\omega}_1^2, \hat{\omega}_2^2, \omega_3^3) = (\omega_1^1, \omega_2^2, \omega_3^3)$, and so on, giving a sequence $\{\hat{\omega}^n\}$ in K . Any limit point ω^* satisfies (B.6). \square

Let $\kappa \in \text{Bel}(\Omega)$ and suppose that the corresponding U satisfies Symmetry and WOI. By Lemma B.4, M_κ can be viewed as a measure on $[\mathcal{K}(S)]^\infty$, and by Lemma B.3, it is exchangeable. Thus we can apply de Finetti’s Theorem (Hewitt and Savage [22]) to M_κ , viewing $\mathcal{K}(S)$ as the one-period state space, to obtain: There exists a unique $\hat{\mu} \in \Delta(\Delta(\mathcal{K}(S)))$ such that

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

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

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

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

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

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

Let $\mu \in \Delta(\text{Bel}(S))$ be the measure induced by $\hat{\mu}$ through the homeomorphism defined in the Choquet Theorem. Apply the Change of Variables Theorem to derive, for any $A \in \Sigma$,

$$\begin{aligned} \kappa(A) &= M_\kappa(\{K \in \mathcal{K}(\Omega) : K \subset A\}) \\ &= \int_{\Delta(\mathcal{K}(S))} \ell^\infty(\{K \in \mathcal{K}(\Omega) : K \subset A\}) d\hat{\mu}(\ell) \\ &= \int_{\text{Bel}(S)} m_v^\infty(\{K \in \mathcal{K}(\Omega) : K \subset A\}) d\mu(v) = \int_{\text{Bel}(S)} v^\infty(A) d\mu(v). \end{aligned}$$

Uniqueness of μ follows from the uniqueness of $\hat{\mu}$ provided by de Finetti’s Theorem. \square

Appendix C. Proofs for Section 3.2

Lemma C.1. *Given a theory, $\kappa(\cdot)$ is well-defined by (1.5)–(1.6).*

Proof. We prove that $\theta \mapsto \nu_\theta^\infty(A)$ is universally measurable for each event $A \subset \Omega$. Then $\kappa = \int_\Theta \nu_\theta^\infty d\mu(\theta)$ is a well-defined capacity.

First, show that G is universally measurable on $\widehat{S} \times \Theta$.²³ Let A be a Borel measurable subset of S . Then, $C \equiv \{K \in \mathcal{K}(S) : K \subset A\}$ is coanalytic by Philippe et al. [34, p. 772]. Observe that

$$\{(\widehat{s}, \theta) : G(\widehat{s} | \theta) \subset A\} = \{(\widehat{s}, \theta) : G(\widehat{s} | \theta) \in C\} = G^{-1}(C)$$

where G is viewed as a function from $\widehat{S} \times \Theta$ to $\mathcal{K}(S)$. Since G is a weakly measurable correspondence, it is a measurable function by Aliprantis and Border [1, Thm. 18.10]. Then, Bertsekas and Shreve [4, Proposition 7.40] imply that $G^{-1}(C) = (\widehat{S} \times \Theta) \setminus G^{-1}(\mathcal{K}(S) \setminus C)$ is coanalytic, and hence universally measurable.

Fix a Borel measurable $A \subset \Omega$. Let φ be the indicator function on $\widehat{S}^\infty \times \Theta$ such that $\varphi(\widehat{s}^\infty, \theta) = 1$ if $G^\infty(\widehat{s}^\infty | \theta) \subset A$, and 0 otherwise. Then, φ is universally measurable.²⁴ Moreover, $\theta \mapsto m_\theta^\infty$ is measurable. Therefore, by Bertsekas and Shreve [4, Proposition 7.46], $\theta \mapsto \nu_\theta^\infty(A) = \int \varphi(\widehat{s}^\infty, \theta) dm_\theta(\widehat{s}^\infty)$ is universally measurable. \square

Proof of Corollary 3.2. By Theorem 3.1, (c) implies (a). By Theorems 3.1 and A.1, (a) implies (b). Clearly (b) implies (c). \square

Appendix D. Proofs for Section 3.4

Denote by $\Psi_n(\cdot)(\omega)$ the empirical frequency measure given the sample ω ; $\Psi_n(A)(\omega)$ is the empirical frequency of the event $A \subset S$ in the first n experiments.

Proof of (3.9). In the entry game, let ν be the belief function on S defined by the probability interval $[0, \eta]$. The LLN in Maccheroni and Marinacci [28] implies that

$$\nu_\eta^\infty(\{\omega \in \Omega : [\liminf \Psi_n(T)(\omega), \limsup \Psi_n(T)(\omega)] \subset [0, \eta]\}) = 1. \tag{D.1}$$

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

$$\begin{aligned} [a > \nu_\eta(B) = 0 \text{ or } b < 1 - \nu_\eta(N) = \eta] &\implies \\ 0 = (\nu_\eta)^\infty(\{[\liminf \Psi_n(T)(\omega), \limsup \Psi_n(T)(\omega)] \subset [a, b]\}) &. \end{aligned} \tag{D.2}$$

Therefore, the representation (3.3) implies that, for every $0 \leq b \leq 1$,

$$\begin{aligned} \mu(\{\eta : 0 \leq \eta \leq b\}) & \\ = U(\{\omega : [\liminf \Psi_n(T)(\omega), \limsup \Psi_n(T)(\omega)] \subset [0, b]\}) & \\ = U(\{\omega : \limsup \Psi_n(T)(\omega) \leq b\}) &. \quad \square \end{aligned} \tag{D.3}$$

²³ A correspondence $\Gamma : \widehat{\Omega} \rightarrow \Omega$ is universally measurable if $\{\widehat{\omega} : \Gamma(\widehat{\omega}) \subset A\}$ is universally measurable for every Borel measurable $A \subset \Omega$.

²⁴ The proof is standard but we show here that φ is universally measurable. First, assume that $B = B_1 \times B_2 \times \dots$ for event $B_i \subset S$. Then,

$$\{(\widehat{s}^\infty, \theta) : G^\infty(\widehat{s}^\infty | \theta) \subset B\} = \bigcap_{i=1}^\infty \{(\widehat{s}_1, \widehat{s}_2, \dots), \theta) : G(\widehat{s}_i | \theta) \subset B_i\}$$

which is universally measurable. As the sets of the form $B_1 \times B_2 \times \dots$ for $B_i \subset S$ generate the Borel σ -algebra on Ω , $\{(\widehat{s}^\infty, \theta) : G^\infty(\widehat{s}^\infty | \theta) \subset B\}$ is universally measurable for every event $B \subset \Omega$.

The following connection between prior beliefs and the certainty equivalents of bets on empirical frequencies, is a corollary of [Theorem 3.1](#).

Corollary D.1. *Adopt the assumptions in [Theorem 3.1](#) and let U and μ be as provided there. Then:*

(a) *For every finite collection $\{A_1, \dots, A_J\}$ of subsets of S , and for all $a_j \leq b_j, j = 1, \dots, J$,*

$$\begin{aligned} &\mu \left(\bigcap_{j=1}^J \{v : [v(A_j), 1 - v(S \setminus A_j)] \subset [a_j, b_j]\} \right) \\ &= U \left(\bigcap_{j=1}^J \{\omega : [\liminf \Psi_n(A_j)(\omega), \limsup \Psi_n(A_j)(\omega)] \subset [a_j, b_j]\} \right). \end{aligned} \tag{D.4}$$

(b) *Let μ' be any probability measure on $Bel(S)$ that agrees with μ on all sets of the form*

$$\{v \in Bel(S) : v(A_1) \geq a_1, \dots, v(A_J) \geq a_J\},$$

where A_j, a_j and J vary over the nonempty subsets of $S, [0, 1]$ and the positive integers respectively. Then $\mu' = \mu$.

Eq. (D.4) relates the prior μ over parameters, here belief functions, to the evaluation of bets on empirical frequencies for the events A_1, \dots, A_J . More precisely, the μ -measures of the sets shown are so related. Part (b) shows that μ is completely determined by its values on these sets.

The proof requires two lemmas. Recall that $\Psi_n(A)(\omega) = \frac{1}{n} \sum_{i=1}^n I(s_i \in A)$ where s_i is the i th component of $\omega \in S^\infty$. 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 D.2. *Let $K \in [\mathcal{K}(S)]^\infty, K = K_1 \times K_2 \times \dots$, and $\alpha \in \mathbb{R}$. Then the following are equivalent:*

- (i) $\liminf_n \Psi_n(A)(\omega) > \alpha$ for every $s_i \in K_i, i = 1, 2, \dots$
- (ii) $\liminf_n \widehat{\Psi}_n(A)(K) > \alpha$.

Proof. (i) \Rightarrow (ii): If $K_i \subset A$, let s_i be any element in K_i , and otherwise, let s_i be any element in $K_i \setminus A$. Then, $I(K_i \subset A) = I(s_i \in A)$ and thus (ii) is implied.

(ii) \Rightarrow (i): If $s_i \in K_i, I(K_i \subset A) \leq I(s_i \in A)$. Thus, if $s_i \in K_i$ for $i = 1, 2, \dots$, then,

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

Lemma D.3. *For any $v \in Bel(S)$:*

- (i) $v^\infty(\{\omega : v(A) < \liminf_n \Psi_n(A)(\omega)\}) = 0$ for each $A \subset S$.
- (ii) $v^\infty(\{\omega : \limsup_n \Psi_n(A)(\omega) < 1 - v(S \setminus A)\}) = 0$ for each $A \subset S$.

Proof. Let $m_\nu \in \Delta(\mathcal{K}(S))$ correspond to ν as in the Choquet Theorem. Fix $A \subset S$. Then,

$$\begin{aligned} & \nu^\infty \left(\left\{ \omega : \nu(A) < \liminf_n \Psi_n(A)(\omega) \right\} \right) \\ &= m_\nu^\infty \left(\left\{ K \in [\mathcal{K}(S)]^\infty : K \subset \left\{ \omega : \nu(A) < \liminf_n \Psi_n(A)(\omega) \right\} \right\} \right) \\ &= m_\nu^\infty \left(\left\{ K \in [\mathcal{K}(S)]^\infty : \liminf_n \widehat{\Psi}_n(A)(K) > \nu(A) \right\} \right) \text{ (by Lemma D.2).} \end{aligned}$$

By the classical LLN, $\widehat{\Psi}_n(A)(K)$ converges to $m_\nu(\{K_1 \in \mathcal{K}(S) : K_1 \subset A\}) = \nu(A)$ m_ν^∞ -almost surely, which implies (i). The proof of (ii) is similar. \square

Proof of Corollary D.1. (a) Because $\nu^\infty(A) = \nu(A)$ for $A \subset S$, $\nu \mapsto \nu(A)$ is universally measurable by Lemma B.1. Hence, every set of the form

$$\{\nu \in \text{Bel}(S) : [\nu(A), 1 - \nu(S \setminus A)] \subset [a, b]\}$$

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

By the LLN in Maccheroni and Marinacci [28], Lemma D.3 and the monotonicity of belief functions,

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

and

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

Moreover, for any belief function γ on Ω , if $\gamma(A) = \gamma(B) = 1$, then $\gamma(A \cap B) = 1$ by the Choquet Theorem (Theorem A.1). Therefore, for the belief function κ in (3.3),

$$\begin{aligned} & \kappa \left(\bigcap_{j=1}^J \left\{ \omega : [\liminf \Psi_n(A_j)(\omega), \limsup \Psi_n(A_j)(\omega)] \subset [a_j, b_j] \right\} \right) \\ &= \int_{\text{Bel}(S)} \nu^\infty \left(\bigcap_{j=1}^J \left\{ \omega : [\liminf \Psi_n(A_j)(\omega), \limsup \Psi_n(A_j)(\omega)] \subset [a_j, b_j] \right\} \right) d\mu(\nu) \\ &= \mu \left(\bigcap_{j=1}^J \left\{ \nu : [\nu(A_j), 1 - \nu(S \setminus A_j)] \subset [a_j, b_j] \right\} \right). \end{aligned}$$

(b) We can identify μ' and μ with measures on $\Delta(\mathcal{K}(S))$. Modulo this identification, we are given that μ' and μ agree on the collection of all subsets of $\Delta(\mathcal{K}(S))$ of the form

$$\bigcap_{j=1}^J \left\{ \ell \in \Delta(\mathcal{K}(S)) : \ell(\{K \in \mathcal{K}(S) : K \subset A_j\}) \geq a_j \right\},$$

for all $J > 0$, $A_j \subset S$ and $a_j \in [0, 1]$. They necessarily agree also on the generated σ -algebra, denoted Σ^* . Therefore, it suffices to show that

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

Step 1. $\ell \mapsto \ell(C)$ is Σ^* -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 $\ell \mapsto \ell(C)$ is Σ^* -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 σ -algebra: (i) $C \in \mathcal{C}$ implies $\mathcal{K}(S) \setminus C \in \mathcal{C}$; (ii) if each $C_j \in \mathcal{C}$, then $\ell \mapsto \ell\left(\bigcup_{j=1}^{\infty} C_j\right)$ is Σ^* -measurable because it equals the pointwise limit of $\ell \mapsto \ell\left(\bigcup_{j=1}^n C_j\right)$ – hence $\bigcup_{j=1}^{\infty} C_j \in \mathcal{C}$.

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

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

Appendix E. Proof of Theorem 4.1

Each belief function ν on $S = \{R, B\}$ corresponds to the probability interval for outcome R given by $[\nu(R), \nu^*(R)]$.

Step 1: Show that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int \int G(\alpha - \Psi_n(\omega)) d\nu^\infty d\mu(\nu) \\ &= \int \min\{G(\alpha - \nu(R)), G(\alpha - \nu^*(R))\} d\mu(\nu). \end{aligned}$$

Without loss of generality, suppose that $\inf_a G(a) = 0$. Define $\bar{G} = \sup_a G(a)$. Since G is quasi-concave, there are two (inverse) functions G_L^{-1} and G_R^{-1} such that

$$\{a : G(a) \geq t\} = [G_L^{-1}(t), G_R^{-1}(t)] \text{ for all } t \in [0, \bar{G}].$$

(The inverses may not be defined at $t = 0$, but this is of no consequence below.) The Dominated Convergence Theorem implies, for any $\nu \in Bel(S)$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int G(\alpha - \Psi_n(\omega)) d\nu^\infty \\ &= \lim_{n \rightarrow \infty} \int_0^{\bar{G}} \nu^\infty(G(\alpha - \Psi_n(\omega)) \geq t) dt \\ &= \lim_{n \rightarrow \infty} \int_0^{\bar{G}} \nu^\infty(G_L^{-1}(t) \leq \alpha - \Psi_n(\omega) \leq G_R^{-1}(t)) dt \\ &= \int_0^{\bar{G}} \lim_{n \rightarrow \infty} \nu^\infty(-G_R^{-1}(t) + \alpha \leq \Psi_n(\omega) \leq -G_L^{-1}(t) + \alpha) dt \end{aligned}$$

$$\begin{aligned}
 &= \int_0^{\bar{G}} I \left[-G_R^{-1}(t) + \alpha \leq v(R) \leq v^*(R) \leq -G_L^{-1}(t) + \alpha \right] dt \\
 &= \int_0^{\bar{G}} I \left[\begin{array}{l} G_L^{-1}(t) \leq \alpha - v^*(R) \leq G_R^{-1}(t), \\ G_L^{-1}(t) \leq \alpha - v(R) \leq G_R^{-1}(t) \end{array} \right] dt \\
 &= \int_0^{\bar{G}} I \left[G(\alpha - v(R)) \geq t, G(\alpha - v^*(R)) \geq t \right] dt \\
 &= \int_0^{\bar{G}} I \left[\min \{ G(\alpha - v(R)), G(\alpha - v^*(R)) \} \geq t \right] dt \\
 &= \min \{ G(\alpha - v(R)), G(\alpha - v^*(R)) \}.
 \end{aligned}$$

Here, $I[\cdot]$ is the indicator function and the fourth equality follows by the LLN in Maccheroni and Marinacci [28]. By the Dominated Convergence Theorem,

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} \int \int G(\alpha - \Psi_n(\omega)) dv^\infty d\mu(v) \\
 &= \int \lim_{n \rightarrow \infty} \int G(\alpha - \Psi_n(\omega)) dv^\infty d\mu(v) \\
 &= \int \min \{ G(\alpha - v(R)), G(\alpha - v^*(R)) \} d\mu(v).
 \end{aligned}$$

Step 2: Show that

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} \operatorname{argmax}_{\alpha \in [0,1]} \int \int G(\alpha - \Psi_n(\omega)) dv^\infty d\mu(v) \\
 &= \operatorname{argmax}_{\alpha \in [0,1]} \lim_{n \rightarrow \infty} \int \int G(\alpha - \Psi_n(\omega)) dv^\infty d\mu(v).
 \end{aligned}$$

There is a unique solution α_n for $\max_{\alpha \in [0,1]} \int \int G(\alpha - \Psi_n(\omega)) dv^\infty d\mu(v)$: Obviously the maximum exists. Uniqueness follows from the strict concavity of $\alpha \mapsto \int \int G(\alpha - \Psi_n(\omega)) dv^\infty$ for each v . Application of the Maximum Theorem completes the proof of this step once we establish the needed continuity, which we do next.

The set $\{1, 2, \dots, \infty\}$ is compact when endowed with the topology generated by singletons $\{n\}$ and sets of the form $\{n, \dots, \infty\}$. Define $F : [0, 1] \times \{1, 2, \dots, \infty\} \rightarrow \mathbb{R}$ by

$$F(\alpha, n) = \begin{cases} \int \int G(\alpha - \Psi_n(\omega)) dv^\infty(\omega) d\mu(v) & n < \infty \\ \lim_{k \rightarrow \infty} \int \int G(\alpha - \Psi_k(\omega)) dv^\infty(\omega) d\mu(v) & n = \infty \end{cases}$$

F is well-defined by Step 1. It is also jointly continuous: We need to check only the case $\alpha_n \rightarrow \alpha$ and $n \rightarrow \infty$. Note that G is uniformly continuous on $[-1, 1]$ and thus that $F(\cdot, n)$ is continuous uniformly in n . Then the desired joint continuity follows from the triangle inequality, that is, from

$$\begin{aligned} & |F(\alpha_n, n) - F(\alpha, \infty)| \\ & \leq |F(\alpha_n, n) - F(\alpha, n)| + |F(\alpha, n) - F(\alpha, \infty)|. \end{aligned}$$

Step 3: Complete the proof. From Steps 1 and 2,

$$\begin{aligned} \alpha_\infty & \equiv \lim_{n \rightarrow \infty} \operatorname{argmax}_\alpha \int \int G(\alpha - \Psi_n(\omega)) d\nu^\infty d\mu(\nu) \\ & = \operatorname{argmax}_\alpha \lim_{n \rightarrow \infty} \int \int G(\alpha - \Psi_n(\omega)) d\nu^\infty d\mu(\nu) \\ & = \operatorname{argmax}_\alpha \int \min \{G(\alpha - \nu(R)), G(\alpha - \nu^*(R))\} d\mu(\nu). \quad \square \end{aligned}$$

References

- [1] C.D. Aliprantis, K.C. Border, *Infinite Dimensional Analysis*, 3rd edition, Springer, 2006.
- [2] N. Al-Najjar, L. de Castro, Parametric representation of preferences, *J. Econ. Theory* 150 (2014) 642–667.
- [3] F.J. Anscombe, R.J. Aumann, A definition of subjective probability, *Ann. Math. Stats.* 34 (1963) 199–205.
- [4] D.P. Bertsekas, S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
- [5] A. Castaldo, F. Maccheroni, M. Marinacci, Random correspondences as bundles of random variables, *Sankhya (Series A)* 80 (2004) 409–427.
- [6] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, L. Montrucchio, Ambiguity and robust statistics, *J. Econ. Theory* 148 (2013) 974–1049.
- [7] F. Ciliberto, E. Tamer, Market structure and multiple equilibria in airline markets, *Econometrica* 77 (2009) 1791–1828.
- [8] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Statist.* 38 (1967) 325–339.
- [9] A.P. Dempster, A generalization of Bayesian inference, *J. Royal Stat. Soc. Series B* 30 (1968) 205–232.
- [10] P. Diaconis, D. Freedman, Partial exchangeability and sufficiency, in: *Proceedings of the Indian Statistical Institute Jubilee Conference on Statistics: Applications and New Directions*, Calcutta, 1981, pp. 205–236.
- [11] J. Dow, S.R. Werlang, Uncertainty aversion, risk aversion, and the optimal choice of portfolio, *Econometrica* 60 (1992) 197–204.
- [12] L.G. Epstein, M. Marinacci, K. Seo, Coarse contingencies and ambiguity, *Theor. Econ.* 2 (2007) 355–394.
- [13] L.G. Epstein, M. Schneider, Learning under ambiguity, *Rev. Ec. Studies* 74 (2007) 1275–1303.
- [14] L.G. Epstein, M. Schneider, Ambiguity, information quality and asset pricing, *J. Finance* 63 (2008) 197–228.
- [15] L.G. Epstein, K. Seo, Symmetry of evidence without evidence of symmetry, *Theor. Econ.* 5 (2010) 313–368.
- [16] B. de Finetti, La prevision: ses lois logiques, ses sources subjectives, *Ann. Inst. H. Poincare* 7 (1937) 1–68. English translation in: H.E. Kyburg, H.E. Smokler (Eds.), *Studies in Subjective Probability*, 2nd edition, Krieger Publishing, Huntington, NY, 1980, pp. 53–118.
- [17] B. de Finetti, On the condition of partial exchangeability, in: *Studies in Inductive Logic and Probability*, vol. 2, University of California Press, Berkeley, 1980, pp. 193–205.
- [18] P. Ghirardato, On independence for non-additive measure, with a Fubini theorem, *J. Econ. Theory* 73 (1997) 261–291.
- [19] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique priors, *J. Math. Econ.* 18 (1989) 141–153.
- [20] F. Gul, W. Pesendorfer, Expected uncertain utility theory, *Econometrica* 82 (2014) 1–39.
- [21] E. Hendon, H.J. Jacobson, B. Sloth, T. Traanaes, The product of capacities and belief functions, *Math. Soc. Sc.* 32 (1996) 95–108.
- [22] E. Hewitt, L.J. Savage, Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.* 80 (1955) 470–501.
- [23] B. Jovanovic, Observable implications of models with multiple equilibria, *Econometrica* 57 (1989) 1431–1437.
- [24] M. Kasy, *Partial Identification, Distributional Preferences, and the Welfare Ranking of Policies*, Harvard, 2014.
- [25] P. Klibanoff, S. Mukerji, K. Seo, Perceived ambiguity and relevant measures, *Econometrica* 82 (2014) 1945–1978.
- [26] D.M. Kreps, *Notes on the Theory of Choice*, Westview, 1988.

- [27] G. Link, Representation theorems of the de Finetti type for (partially) symmetric probability measures, in: *Studies in Inductive Logic and Probability*, vol. 2, University of California Press, Berkeley, 1980, pp. 207–231.
- [28] F. Maccheroni, M. Marinacci, A strong law of large numbers for capacities, *Ann. Prob.* 33 (2005) 1171–1178.
- [29] C.F. Manski, Choosing treatment policies under ambiguity, *Annual Rev. Econ.* 3 (2011) 25–49.
- [30] C.F. Manski, Status quo deference and policy choice under ambiguity, *J. Institutional and Theor. Econom.* 169 (2013) 116–128.
- [31] C.F. Manski, Diagnostic testing and treatment under ambiguity: using decision analysis to inform clinical practice, *Proceedings of the National Academy of Sciences* 110 (2013) 2064–2069.
- [32] M. Marinacci, Learning from ambiguous urns, *Statistical Papers* 43 (2002) 145–151.
- [33] I. Molchanov, *Theory of Random Sets*, Springer, 2005.
- [34] F. Philippe, G. Debs, J.Y. Jaffray, Decision making with monotone lower probabilities of infinite order, *Math. Oper. Res.* 24 (1999) 767–784.
- [35] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57 (1989) 571–587.
- [36] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [37] G. Shafer, Belief functions and parametric models (with commentary), *J. Royal Stat. Soc. Series B* 44 (1982) 322–352.
- [38] M. Talagrand, Capacités invariantes extrémales, *Ann. Inst. Fourier (Grenoble)* 28 (1978) 79–146.
- [39] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010) 167–195.
- [40] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003) 147–165.
- [41] P.P. Wakker, Dempster belief functions are based on the principle of complete ignorance, *Intern. J. Uncert. Fuzziness and Knowledge-Based Systems* 8 (2000) 271–284.