Bayesian Inference and Non-Bayesian Prediction and Choice: Foundations and an Application to Entry Games with Multiple Equilibria*

Larry G. Epstein Kyoungwon Seo

February 27, 2013

Abstract

As a central motivating example, we consider a policy maker facing a cross-section of markets in which firms play an entry game. Her theory is Nash equilibrium and it is incomplete because there are multiple equilibria and she does not understand how equilibria are selected. This leads to partial identification of parameters when drawing inferences from realized outcomes in some markets and to ambiguity when considering (policy) decisions for other markets. We model both her inference and choice. The central component of the model is a generalization of de Finetti's exchangeable Bayesian model to accommodate ambiguity.

*Boston University, lepstein@bu.edu. We gratefully acknowledge the financial support of the National Science Foundation (awards SES-0917740 and SES-0918248) and discussions with Soo Hong Chew, Marc Henry, Hiro Kaido, Peter Klibanoff and Minjae Song. We also thank three referees, and seminar participants at Harvard, Warwick, Northwestern, Austin, Yale, Cal Tech and Wisconsin. Some of the material in this paper was formerly contained in the paper "A de Finetti Theorem for Capacities," originally circulated in 2009, which no longer exists as a separate paper.

1. Introduction

1.1. Motivation - Entry Games

Consider a policy maker (PM) who must choose a policy that will affect a number of markets. The consequences of the policy depend on firm behavior (for example, entry decisions) which is uncertain. However, the PM has data on behavior in related markets. Thus she wants to learn from these data and then choose a policy. Our goal is to model both how she does inference and how she chooses, and to do so in a way that respects her theory of the environment (how data are generated and how firms behave) and also her concerns. We illustrate her theory and concerns next.

There are I markets, with two firms in each market. The PM assumes that in the i^{th} market, firms $j = 1, 2$ play the entry game described by the payoff matrix shown.[1]

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

The parameter η lies in $(0, 1]$ and the ϵ_{ij} 's are observed by players but not by the PM. She views η as common across markets and the ϵ_{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, 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

$\{B, N\}$	if $0 \leq \epsilon_{i1}, \epsilon_{i2} \leq \eta^{1/2}$	(1.1)
$\{N\}$	otherwise,	

where B denotes the outcome where both firms enter and N the outcome where neither firm enters.

[1] This entry game, taken from Jovanovic (1989), serves as a running example. It illustrates the important features of a class of models used in the applied IO literature, including, for example, in Bresnahan and Reis (1991), Berry (1992), Tamer (2003), Ardillas-Lopez and Tamer (2008), Ciliberto and Tamer (2009) and Bajari-Hong-Ryan (2010). Many assumptions 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 ϵ's. Section 5 shows how more general entry games can be accommodated.

Ignorance of the selection mechanism creates difficulties for both inference and choice. Because η is assumed to be common across markets, one can hope to learn about η from outcomes observed in some markets. However, inference is complicated by the fact that a given sample can be interpreted in different ways: an observed outcome N could be due to the ϵ's and η satisfying the inequality indicated in (1.1) and N being selected, or it could reflect violation of the inequality and N being the unique equilibrium. Consequently, one may not be able to point identify η even with an infinite set of data. In general, η is identified only up to an interval, that is, it is partially identified.[2] The multiplicity of equilibria also makes (prediction and) choice more difficult, and this is so even if the value of η is known. For example, being agnostic about the selection mechanism suggests that the individual may not be willing to make a probabilistic prediction about the outcome in the next market. Without taking a stand on selection, one can say only that the probability of B lies in the interval $[0, \eta]$.

Consider now a PM in the above context and assume also that she has observed the entry outcomes in a subset of markets. We argue in Section 1.3 that the literature provides limited guidance as to how she might proceed. Our objective is to describe a way for her to go about both the inference and choice components of her problem. Our approach is axiomatic so that assumptions are explicit and arguably simple and she can judge if they suit her.

We take the core issue to be ignorance about selection and the desire to make decisions that are robust to this limitation. To elaborate, imagine the PM having the following perspective. She believes that a complete theory of 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 B 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.* Finally, with regard to inference, though the PM tries to learn about η, because she cannot even formulate a theory of selection she cannot learn about selection. There is an analogy with learning from a sequence of Ellsberg urns, each with 100 balls that are labelled either B or N. Suppose that there is repeated sampling from consecutive urns and that the individual has the following perception of the urns: the fraction α of balls has a composition that is common across all urns, while

[2] Tamer (2010) surveys the econometric literature on partial identification. See below for more references.

the remaining component varies across urns in a way that is not understood at all. The individual can then hope to learn about the common component but plausibly does not even attempt to learn about the others.

An important ingredient of the entry game is the a priori dichotomy between two kinds of uncertainty: the PM is confident enough to assign probabilities to the ϵ's, but she does not understand selection well enough to posit a probabilistic selection mechanism. A general model would permit ambiguity about both. We adopt the dichotomy emphasized in the cited applied literature and propose solutions to modeling problems that have not been adequately addressed even for this special case.

The paper proceeds as follows. In the remainder of this introduction we outline our model and then explain the value-added relative to existing literature. Section 2 presents the axiomatic foundations of our model of choice and inference. The implied representations of utility and updating are described in Section 3, specifically in Theorem 3.1 which is our main result. Section 4 demonstrates tractability of the model by considering the problem of optimally predicting empirical frequencies within the entry game. The concluding section outlines an extension to more general entry games and considers additional related literature. In particular, the connection between axiomatic decision theory and statistical decision theory has been emphasized also by Stoye (2012) and thus we compare our approach with Stoye's. Proofs and technical details are collected in appendices.

1.2. Model outline

We outline our model of choice and inference here in the context of the Jovanovic entry game to communicate the essential points. See Section 3 for the general model, including all technical details, and Section 2 for axiomatic foundations.

As in our axiomatic characterization, take $I = \infty$. The set of all possible sequences of market outcomes, or the full state space, is

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

where $S = \{B, N\}$. Frequently, we write B_i and N_i to indicate elements in i-th copy of S. A policy maps a sequence of market outcomes into payoffs, assumed to lie in the unit interval. Therefore, to model policy choice, consider the set $\mathcal{F}$ of all acts $f : \Omega \to [0, 1]$, and a binary relation $\succeq$ on $\mathcal{F}$, interpreted as the ex ante (prior to sampling) preference order. We take $\succeq$ to be ambiguity averse and to be a special case of both the multiple-priors model (Gilboa and Schmeidler

(1989)) and Choquet expected utility (Schmeidler (1989)); we call the special case *belief function utility*.[3] The connection to ambiguity aversion is suggested by the description of the entry game in terms of 'ignorance' of the selection mechanism and 'two kinds of uncertainty' which bring to mind the distinction between risk and ambiguity.

To define belief function utility, consider first acts that depend only on the outcome in market i. Then uncertainty is represented by the state space $S = \{B, N\}$, and an issue is how to arrive at beliefs on S in light of ignorance of selection. We describe one way to do so. Each given parameter η induces, via (1.1), the *equilibrium correspondence* Γ_η,

$$\Gamma_\eta : \{\epsilon_i = (\epsilon_{i1}, \epsilon_{i2})\} = [0,1]^2 \rightsquigarrow \{B, N\}.$$

The PM holds the uniform distribution, denoted m, on $[0,1]^2$. Therefore, a conservative attitude leads to beliefs on S represented by θ_η, where, for any $A \subset S$,

$$\theta_\eta(A) = m\left(\left\{\epsilon_i \in [0,1]^2 : \Gamma_\eta(\epsilon_i) \subset A\right\}\right). \tag{1.2}$$

Then θ_η is a *belief function on* S, and it admits the explicit representation

$$\theta_\eta(B) = 0,\ \theta_\eta(N) = 1 - \eta, \text{ and } \theta_\eta(\{B, N\}) = 1. \tag{1.3}$$

Evidently, θ_η is nonadditive and is a special case of a capacity as defined by Schmeidler (1989).[4] Each θ_η is associated with a set of priors via its core, *core* (θ_η), which is the set of all probability measures P on $\{B, N\}$ that dominate θ_η, that is,

$$core\,(\theta_\eta) = \{P \in \Delta(S) : P(A) \geq \theta_\eta(A), \text{ for all } A \subset S\}. \tag{1.4}$$

Then *core* (θ_η) consists of all measures for which the probability of B lies in the interval $[0, \eta]$.

[3] Appendix A provides some background material regarding belief functions and belief function utility.

[4] Like many other papers in the literature, see Berry (1992) and Ciliberto and Tamer (2009), for example, we consider only pure strategies throughout. Arguably, pure strategies are more intuitive and computationally simpler in applications. In addition, mixed strategy equilibria do not induce a capacity in general because the convex hull of the set of mixed strategy equilibria is not necessarily the core of a capacity. Mixed strategy equilibria are accommodated by the multiple-priors model of preference used in Epstein and Seo (2010).

Next define beliefs on Ω using a similar algorithm. Each parameter η induces the *equilibrium sequence correspondence* Γ_η^∞,

$$\Gamma_\eta^\infty : [0,1]^2 \times [0,1]^2 \times ... \rightsquigarrow \{B,N\}^\infty,$$

where

$$\Gamma_\eta^\infty (\epsilon_1, \epsilon_2, ...) \equiv \Gamma_\eta (\epsilon_1) \times \Gamma_\eta (\epsilon_2) \times ...$$

We are given that the PM views the ϵ_i's as being i.i.d. across markets. Therefore, she employs the i.i.d product measure m^∞ on $[0,1]^2 \times [0,1]^2 \times ...$ Accordingly, a conservative attitude leads to beliefs on Ω represented by the capacity $(\theta_\eta)^\infty$, where, for any $A \subset \Omega$,

$$(\theta_\eta)^\infty (A) = m^\infty \left(\left\{ (\epsilon_i) \in [0,1]^2 \times [0,1]^2 \times ... : \Gamma_\eta^\infty((\epsilon_i)) \subset A \right\} \right). \tag{1.5}$$

Then $(\theta_\eta)^\infty$ is a belief function on Ω. We refer to it as the *i.i.d. product of* θ_η.

The preceding takes η as given, hence known, though the PM may be uncertain about its value. Accordingly, suppose that she forms a prior μ over possible values of η, (thereby excluding ambiguity about η), and that her beliefs about Ω are given by the 'average' belief function ν_μ:

$$\nu_\mu (\cdot) = \int_{(0,1]} (\theta_\eta)^\infty (\cdot)\, d\mu (\eta). \tag{1.6}$$

Finally, her utility function is given by the Choquet expected utility function:[5] for every f in $\mathcal{F}$,

$$U(f) = \int_\Omega f d\nu_\mu. \tag{1.7}$$

(This is a special case of what we later call belief function utility.) Note that the utility function is completely specified by the measure μ.

One way to see that this utility function captures a concern with unknown heterogeneity and correlation across markets is by examining the core of each i.i.d. product $(\theta_\eta)^\infty$.[6] For P a measure on Ω, denote by $mrg_{\{1,2\}}P$ the marginal

[5] The connection to the multiple-priors model is explained in Appendix A. Integration in (1.7) is in the sense of Choquet. The integrand is f rather than $u \circ f$ for some felicity function u, because outcomes are denominated in utils.

[6] $core\left((\theta_\eta)^\infty\right)$ is defined by the counterpart of (1.4). The argument here is based solely on the functional form and hence is only suggestive. Section 2 provides a behavioral argument.

on $S_1 \times S_2$. Then[7]

$$\begin{aligned}\{mrg_{\{1,2\}}P &: P \in core\left((\theta_\eta)^\infty\right)\} = \\ &(1-\eta)^2 \Delta\left(\{N_1\} \times \{N_2\}\right) + \eta^2 \Delta\left(\{B_1, N_1\} \times \{B_2, N_2\}\right) \\ &\eta(1-\eta)\left[\Delta\left(\{B_1, N_1\} \times \{N_2\}\right) + \Delta\left(\{N_1\} \times \{B_2, N_2\}\right)\right].\end{aligned} \tag{1.8}$$

This set of probability mixtures attaches weight η^2 to the set of all probability measures on $\{B_1, N_1\} \times \{B_2, N_2\}$, including both nonproduct measures and non-identical products, thus indicating that uncertainty about both correlation and heterogeneity is reflected in beliefs. (We will see later that there is aversion to this uncertainty.)

To discuss inference, consider as a primitive not only ex ante preference, denoted now $\succeq_0$, but also the conditional preference order $\succeq_{n,s^n}$ that prevails after observing the sample $s^n = (s_1, ..., s_n)$ of outcomes in markets 1 through n. Assume that the corresponding utility function U_{n,s^n} is also a belief function utility, that is, it has the form in (1.7) though with the posterior beliefs μ_{n,s^n} about η. Then the inference problem consists of how the posterior μ_{n,s^n} is related to the prior μ_0 and the sample s^n. This is a matter of updating probability measures, but it is not a standard updating problem because of the difficulty mentioned in the previous section regarding how to interpret a signal. Because each parameter value η is associated with the probability interval $[0, \eta]$ for observing B, there is a *set of likelihoods* and thus Bayes' rule does not apply immediately. The model dictates that the PM average over these likelihoods, specifically that she update μ_0 by Bayes' rule using a likelihood $L(\cdot \mid \eta)$ of the form

$$L(\cdot \mid \eta) = \int_{q \in \Delta(\{B,N\})} q^\infty(\cdot)\, d\lambda_\eta(q), \tag{1.9}$$

where each λ_η is a (subjective) probability measure over selection mechanisms (represented by q). Thus the complete model of choice and inference is specified by μ_0 and the collection $\{\lambda_\eta\}_{\eta \in (0,1]}$.

Similar likelihood specifications have been used in the literature. Acemoglu et al. (2009) use such a likelihood to model a difficult to interpret signal. We interpret the likelihood similarly: it is as if the individual is uncertain, to a degree

[7] This expression follows from (A.8) in Appendix A. It is readily understood informally. For example, note that the weight η^2 attached to $\Delta(\{B_1, N_1\} \times \{B_2, N_2\})$ is the probability according to the product measure m^2 that (ϵ_1, ϵ_2) lies in the region for which there are two equilibria in both markets 1 and 2.

represented by λ_η, what any given realized sample reveals about η. Accordingly, the subjective nature of the λ_η's is natural. Some PM's may update as if each λ_η is uniform over $[0, \eta]$. (Here, because each q in $\Delta(\{B, N\})$ can be identified with the probability of B, we identify each λ_η with a measure on the unit interval.) Others may process the signal provided by an observed market outcome B as if its likelihood were unambiguously equal to $\eta/2$.[8]

In the more directly relevant literature, Moon and Schorfheide (2012) use a likelihood function of this form in their Bayesian econometric approach to inference in partially identified models. Though one can invoke the subjective expected utility framework for foundations to (1.9), we argue below that such a model of choice does not capture an aversion to unknown heterogeneity and correlation. A possibly surprising feature of our model is the demonstration that Bayesian inference is compatible with non-Bayesian choice and the noted aversion.

1.3. What does the literature provide?

One way to model the entry game is to use the *exchangeable Bayesian model*. Savage (1972) and Anscombe and Aumann (1963) axiomatize the subjective expected utility model of choice and de Finetti (1937) shows that exchangeability (or symmetry), the property that the probability of any finite set of outcomes does not depend on the order in which the outcomes are realized, characterizes the "conditionally i.i.d." form for the predictive prior $P \in \Delta(\Omega)$:

$$P(\cdot) = \int_{\Delta(S)} q^\infty(\cdot)\, d\mu(q). \tag{1.10}$$

Here μ represents beliefs (either a prior or posterior after updating) on $\Delta(S)$. Updating is done by application of Bayes' rule.

Expressed in terms of behavior, exchangeability is the assumption that there is indifference between any two bets that differ only by a permutation or reordering of markets; such indifference is natural in the Jovanovic entry game because there is no reason given to distinguish between markets. However, we argue in Section 2 that the Independence Axiom for preference implies indifference also to uncertainty about heterogeneity and correlation. The de Finetti representation is strongly suggestive: though the beliefs in (1.10) reflect uncertainty about the

[8] They behave differently. Only the latter type would converge, in a given sample, to certainty that η equals twice the limiting empirical frequency of B.

true selection probability, they also indicate certainty that the selection mechanism is i.i.d. across markets, thus excluding any concern about heterogeneity and correlation. This motivates us to generalize the Bayesian model, while retaining exchangeability, that is, the noted indifference to the ordering of markets.

There is a large literature on partially identified models. We share the view underlying much of that literature, and that is expressed most forcefully by Manski (2003) and Tamer (2003, 2010), that modelers should avoid assumptions that are driven by convenience (for example, adding an ad hoc assumption about the selection mechanism in order to permit point indentification of η) rather than by economic theory.[9] These authors often have in mind an empirical modeler, but the same principle has merit when applied to a decision maker such as our PM. Our emphasis on robustness to model uncertainty, specifically to the i.i.d. assumption across markets, can be understood in terms of the following slight restatement of Manski's (2003) "law of decreasing credibility": the credibility of inference and (we would add) the desirability of choice decrease with the strength of the assumptions maintained.

We differ from the partial identification literature in the specifics of how and in what sense robustness is achieved or modeled. Most of the literature studies inference and estimation using a frequentist approach.[10] These papers do not explicitly address choice in their formal analyses. However, our presumption is that, as stated by Tamer (2010, p. 174):

> *One main motivation for empirical work in economics is to evaluate policies, with an important purpose of decision making*

Therefore, we ask how this frequentist based literature feeds into modeling choice. With unlimited data, one can in principle identify a set of values for η, which in turn yields a set of predictive measures. These might be used as in the multiple-priors model to guide choice. With finite samples, one has only an estimate of the identified set and thus the added layer of uncertainty due to estimation error. It is not clear to us how to base decision making on any given estimation results. The difficulty, it seems to us, is inherent in the approach of first doing inference/estimation, and only afterwards worrying about how to use the output to make decisions. Rather than viewing inference and choice as separable in this way, we take seriously that decision making is the ultimate goal. Accordingly, in

[9] See also Berry and Tamer (2006) and Echenique and Komunjer (2009).

[10] Chernozhukov et al. (2007) is a prominent example.

our model choice drives inference in the sense that preference is the primitive and implications for inference are derived from assumptions about preference.

Bayesian statistical methods have also been applied to partially identified models (Moon and Schorfheide (2012) and Liao and Jiang (2010), for example). Though these authors are not explicit about how to model choice, presumably they have in mind subjective expected utility maximization. As mentioned, we argue in this paper that it does not capture the story surrounding the entry game.

Stoye (2012), Kitagawa (2012) and Menzel (2011) study robust approaches to statistical decision problems for partially identified models. They do not address robustness with respect to correlation and heterogeneity; nor do they deal with policy choice. Section 5.3 elaborates on differences between studies focussed on statistical decision making and our approach which is focussed on, and driven by, policy choice.

The only papers of which we are aware that explicitly address policy choice in the context of partially identified models are Manski (2011, 2012, 2013) and Kasy (2012).[11] Their approaches are nonaxiomatic and their models are much different than ours.

Finally, in this introduction we elaborate on how we see our axioms and the value of our axiomatic approach. For the most part, we do not see our model as descriptive in the sense of explaining observed behavior of policy makers, for example. (The qualification is added because the model does 'explain' the Moon and Schorfheide (2012) statistical procedure.) Neither is the model normative in the strong sense that the Savage axioms are often seen. We do not claim that our axioms would be acceptable to everyone.[12] However, we do suggest that they lay out in simple and explicit terms the principles characterizing our model, which could be useful to the PM we have described: she does not understand the selection mechanism and she seeks a course of action that is robust to the implied uncertainty about heterogeneity and correlation across markets. She feels uncomfortable or dissatisfied with existing models of choice and needs some guidance. Therefore, even if she does not find our axioms compelling, she might very well find them to be sensible and helpful; at least she will be able to judge.

[11] Ciliberto and Tamer (2009) conduct a counterfactual analysis of policy change in airline markets, though they do not take the natural next step of modeling the choice of policy.

[12] Note that even the Savage axioms are not always seen as compelling. For example, Gilboa et al (2012) see the Ellsberg paradox as a normative critique of expected utility theory, which is a viewpoint that we share.

2. Foundations

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

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

Objects of choice are (Borel measurable and simple, that is, finite-ranged) acts $f : \Omega \to [0, 1]$. The set of all acts is $\mathcal{F}$. Binary acts are called bets. The bet that pays 1 util if there are two entrants in the first market and none in the second is denoted B_1N_2. The bet (with payoffs 1 and 0) that the first two markets have the same number of entrants is denoted $\{B_1B_2, N_1N_2\}$. Similarly for other bets.

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-Aumann 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 choice of acts both ex ante and after observing the outcomes of n experiments, where n is arbitrary. We emphasize that, for reasons given below, we assume that one sample only is observed and that therefore, updating is done only once, as opposed to repeatedly with gradually increasing sample size. (In the entry game example, the PM observes outcomes in some markets and then chooses a policy that affects remaining markets. Further revisions are not modeled.) To model both choice and inference, we adopt as primitives the set of conditional preferences $\{\succeq_{n,s^n} : n \geq 0, s^n \in S^n\}$, where $\succeq_{n,s^n}$ is the preference on $\mathcal{F}$ conditional on having observed the outcomes $s^n = (s_1, ..., s_n)$ in the first n experiments; $\succeq_0$, corresponding to $n = 0$, denotes ex ante preference. We specify a number of axioms for $\{\succeq_{n,s^n}\}$.

The ordering of experiments is not temporal, nor is it important. One should think of a cross-sectional setup, where the n experiments producing the sample s^n were carried out simultaneously and all remaining experiments will be conducted simultaneously. It is convenient for the formalism to fix an order, which we do, but it is arbitrary.

To state the first axiom, we must define "belief function utility". First we generalize the definition of belief function from (1.5) and the entry game example.[13] Refer to (2.1). The objects of choice are acts f over Ω, and thus beliefs ν on Ω are important. To formulate them, the decision maker employs an auxiliary

$$\begin{array}{ccc} (\widehat{\Omega}, m) & \overset{\Gamma}{\rightsquigarrow} & (\Omega, \nu) \\ & & \downarrow_f \\ & & [0,1] \end{array} \tag{2.1}$$

state space $\widehat{\Omega}$ where her understanding permits her to assign probabilities using a measure m. The auxiliary space provides a coarse picture of Ω in that each point in $\widehat{\Omega}$ corresponds to a set of points in Ω, that is, there is a correspondence Γ from $\widehat{\Omega}$ into Ω. Awareness of this coarseness and a conservative attitude lead to beliefs on Ω represented by ν, where

$$\nu\left(A\right) = m\left(\left\{\widehat{\omega} \in \widehat{\Omega} : \Gamma(\widehat{\omega}) \subset A\right\}\right). \tag{2.2}$$

Any function ν on the Borel σ-algebra of Ω that can be constructed in this way is called a *belief function* on Ω.[14] Refer to $(\widehat{\Omega}, m, \Gamma)$ as representing, or generating, ν. In the entry game example, the belief function defined via (1.5) has $\widehat{\Omega} = \Pi_{i=1}^{\infty} [0,1]^2$, m equal to the i.i.d. product of the uniform measure and Γ equal to the equilibrium sequence correspondence.

A function $U : \mathcal{F} \to \mathbb{R}$ is called a *belief function utility* (for the state space Ω) if there exists a belief function ν on Ω such that

$$U\left(f\right) = U_{\nu}\left(f\right) = \int_{\Omega} f d\nu, \text{ for all } f \text{ in } \mathcal{F}. \tag{2.3}$$

Here integration is in the sense of Choquet and thus every belief function utility is a special case of Choquet expected utility (Schmeidler (1989)).

For all the axioms that follow, the quantifier "for all n and s^n" should be understood.

[13] See Appendix A for more on belief functions and the corresponding utility functions.

[14] More precisely, we take $\widehat{\Omega}$ to be compact metric, m a Borel probability measure, the correspondence Γ measurable and nonempty-compact-valued, and ν is defined on the Borel σ-algebra of $\widehat{\Omega}$. Equivalent definitions of belief functions are described in Appendix A. If Γ is singleton-valued and hence a random variable, then ν is a probability measure and (2.2) is the familiar formula for computing induced distributions.

Axiom 1 (Belief Function Utility). *Every preference $\succeq_{n,s^n}$ 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 et al. (2007) and Gul and Pesendorfer (2010) describe behavioral foundations for (2.3). 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 subjective 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 that we call Symmetry. Let Π be the set of (finite) permutations on $\mathbb{N}$. For $\pi \in \Pi$ and $\omega = (s_1, s_2, ...) \in \Omega$, let $\pi\omega = \left(s_{\pi(1)}, s_{\pi(2)}, ...\right)$. Given an act f, define the permuted act πf by $(\pi f)(s_1, ..., s_n, ...) = f\left(s_{\pi(1)}, ..., s_{\pi(n)}, ...\right)$. For example, if $f = B_1 N_2$ and π switches 1 and 2, then $\pi f = N_1 B_2$. An act is said to be *finitely-based* if it depends on the outcomes of only finitely many experiments.

Axiom 2 (Symmetry). *For all finitely-based acts f and permutations π,*

$$f \sim_{n,s^n} \pi f.$$

Symmetry is intuitive in situations where the temporal ordering of experiments is not important and information about the experiments is symmetric. This intuition applies also to conditional preference, even if the sample exhibits the alternating pattern $B_1, N_1, B_2, N_2, ..., B_{n/2}, N_{n/2}$. The outcomes of experiments 1 to n constitute cross-sectional data and their ordering has no significance.

Note that symmetry of information does not imply that information is substantial; in fact there could be no information available at all about any of the experiments and about how they are related. Thus Symmetry 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 acts f and g as mutually *orthogonal* if they depend on disjoint sets of experiments; write $f \perp g$. Our main axiom is:

Axiom 3 (Weak Orthogonal Independence (WOI)). *For all* $0 < \alpha \leq 1$, *and all finitely based acts* f', f *and* g *such that* $f' \perp g$ *and* $f \perp g$,

$$f' \succeq_{n,s^n} f \iff \alpha f' + (1-\alpha)\, g \succeq_{n,s^n} \alpha f + (1-\alpha)\, g.$$

To minimize notational clutter, refer to a generic preference $\succeq$ satisfying the axiom. Given also Belief Function Utility, then WOI is satisfied if and only if the corresponding utility function U satisfies:[15] For all α and finitely-based and orthogonal acts f and g,

$$U(\alpha f + (1-\alpha)\, g) = \alpha U(f) + (1-\alpha)\, U(g). \tag{2.4}$$

We use this characterization of WOI frequently in the sequel. Note that the belief function utility U provided by our first axiom provides a certainty equivalent because, by (2.3), any act f is indifferent to the constant act giving $U(f)$ in every state. Therefore, the expression (2.4) is a meaningful statement about preference.

Turn to interpretation of the axiom. The Independence axiom requires the similar invariance of rankings for all (not necessarily orthogonal) 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 are germane to the entry game and that are excluded by Independence but permitted by WOI. The first is simply ambiguity about the outcome in any *single* market. Even given knowledge of η, ignorance of the selection mechanism suggests the perception that the probability of B could lie anywhere in $[0, \eta]$. Given that no entry can be a unique equilibrium but that dual entry cannot, the strict ranking $N_1 \succ B_1$ is intuitive. Without loss of generality, suppose that $.8N_1 \sim B_1$, that is, indifference is restored by suitably reducing the winning prize when there is no entry. Then the intuitive ranking familiar from the 2-color Ellsberg Paradox is that

$$\tfrac{1}{2}B_1 + \tfrac{1}{2}(.8N_1) \succ B_1, \tag{2.5}$$

which contradicts Independence. Gilboa and Schmeidler (1989) describe the value of such randomization as due to its smoothing out ambiguity, or, adapting finance terminology, because the bets being mixed may "hedge" one another.

---

[15] Assume WOI and let f' be constant at level $U(f)$, so that $f' \sim f$. Because constant acts are orthogonal to every act, deduce that $\alpha U(f) + (1-\alpha)\, g \sim \alpha f + (1-\alpha)\, g$ and thus $\alpha U(f) + (1-\alpha)\, U(g) = U(\alpha U(f) + (1-\alpha)\, g) = U(\alpha f + (1-\alpha)\, g)$. The first equality is due to (A.6).

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

$$\tfrac{1}{2}B_1 + \tfrac{1}{2}N_1 \succ \tfrac{1}{2}B_1 + \tfrac{1}{2}N_2. \tag{2.6}$$

The act on the left perfectly hedges uncertainty about the first experiment and yields $\frac{1}{2}$ with certainty. But the act on the right also involves uncertainty about possible differences in the selection mechanism across markets. For example, if selection favors dual entry in the first market and no entry in the second, that is a good scenario for $\frac{1}{2}B_1 + \frac{1}{2}N_2$. However, under the reverse scenario, the 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, ambiguous heterogeneity suggests (2.6).

Finally, we illustrate behavior that reveals a concern with correlation, which we take to mean roughly a concern that the selection mechanism may follow some unknown "patterns." Consider betting that the outcomes are identical in the first two markets versus betting that they are identical in the first and third markets. Symmetry implies indifference. However, there is intuition for the following rankings contradicting Independence:

$$\begin{aligned} & \tfrac{1}{2}\{B_1B_2, N_1N_2\} + \tfrac{1}{2}\{B_1B_3, N_1N_3\} \\ \succ\ & \{B_1B_2, N_1N_2\} \sim \{B_1B_3, N_1N_3\}. \end{aligned} \tag{2.7}$$

Suppose the probability of selection of B depends positively on an unknown 'omitted' variable. The variable may be similar in markets one and two (and differ between markets one and three), which would favor the bet $\{B_1B_2, N_1N_2\}$, or the variable may be similar in markets one and three (and differ in markets two and three), which would favor the other bet. Which is the case is uncertain. The mixture is strictly preferable because it smooths out this uncertainty.

It is comforting that WOI permits (2.5)-(2.7), but it remains to determine what kind of behavior it excludes. Interpret the axiom in the entry game setting. Because it imposes that bets on outcomes in different markets do not hedge one another, roughly speaking the assumption is that disjoint sets of markets perceived to be 'not connected.' One connection is that market outcomes depend on the common factor η. If there is ambiguity about η, then the sort of hedging gains pointed to by Gilboa and Schmeidler (1989) would lead to violations of WOI. Thus the axiom excludes ambiguity about the parameter η. In addition, even given η, it excludes the perception that markets are "stochastically dependent." For example, it excludes certainty that selection is identical in all markets, whether it be that B is always selected, or alternatively that N is selected in all markets.[16] In that case, one would expect that (with rescaling as in (2.5))

$$\tfrac{1}{2}B_1 + \tfrac{1}{2}(.8N_2) \succ B_1 \sim .8N_2,$$

contrary to (2.4).

One would like to say more about the kind of (conditional) stochastic independence being assumed via WOI, especially because while referring to stochastic independence we also argue that ambiguous correlation is accommodated by the model. The difficulty is that "stochastic independence" is multifaceted if there is ambiguity and not well understood behaviorally (see Ghirardato (1997), for example). However, one can view the axiom as providing behavioral meaning for one form of (conditional) stochastic independence.

A major appeal of the axiom in the present context is its simplicity, which promises that a decision-maker would be able to understand it and either accept or reject the axiom. In the entry game, indifferences required by WOI, such as

$$\tfrac{1}{2}B_1 + \tfrac{1}{2}N_3 \sim \tfrac{1}{2}B_2 + \tfrac{1}{2}N_3,$$

seem intuitive where selection is poorly understood, and in any case are simple enough that a decision-maker would be able to agree or not.

The final three axioms are focussed on conditioning.

Axiom 4 (Consequentialism). *For all acts f' and f, $f' \sim_{n,s^n} f$ if $f'(s^n,\cdot) = f(s^n,\cdot)$.*

This familar axiom states that when evaluating acts conditional on the node (n, s^n), the decision maker cares only about what these acts promise on the continuation from that node: unrealized parts of the tree do not matter.

Axiom 5 (Commutativity). *For all permutations π, $\succeq_{n,\pi s^n} = \succeq_{n,s^n}$*

[16] To be perfectly clear, our intention here is to certainty that selection is perfectly correlated, which excludes the possibility that other forms of correlation exist. The axiom and model admit perfect correlation as a possibility in the mind of the decision maker. It is the exclusion of all other correlation patterns on her part that is contradicted by the axiom.

There is no natural ordering of cross-sectional data and thus the order of past observations should not matter.

If $k < n$, write $s^n = \left(s^k, s^{n\backslash k}\right)$, where $s^{n\backslash k}$ denotes $(s_{k+1}, ..., s_n)$

Axiom 6 (Weak Dynamic Consistency (WDC)). *For any $0 \leq k < n$, sample s^k and acts f', f over $S_{n+1} \times S_{n+2} \times ...$,*

$$\text{if } f' \succeq_{n,(s^k,s^{n\backslash k})} f \text{ for all } s^{n\backslash k}, \text{ then } f' \succeq_{k,s^k} f,$$

$$\text{and, if in addition } f' \succ_{n,(s^k,s^{n\backslash k})} f \text{ for some } s^{n\backslash k}, \text{ then } f' \succ_{k,s^k} f.$$

Suppose first that $k = 0$. Let f' and f describe two policies that pertain only to markets $n+1$ and beyond. Data are to be collected describing the outcomes, B or N, realized in the first n markets. Suppose that for every possible realization of the sample, f' would be preferred to f ex post. Then the axiom requires that also ex ante, prior to collecting the data, f' be preferred to f. (Further, strict preference conditional on at least one sample implies strict preference ex ante.) The axiom weakens the usual dynamic consistency assumption, which would have the identical statement except that f' and f would be allowed to vary over all acts in $\mathcal{F}$. The weakening refers only to 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).

The intuition for such (weak) consistency between post sample preferences and ex ante preference is not due to a special status for the ex ante stage. Thus the axiom extends it to also to the 'interim stage' given by k and s^k.

Given the prescriptive nature of our model and the strong normative appeal of dynamic consistency, why do we adopt only the weaker axiom? The reason is that dynamic consistency conflicts with other desiderata. Assume that preferences satisfy Symmetry, Consequentialism and very weak continuity and monotonicity properties as in Epstein and Seo (2011); the latter two properties are jointly much weaker than Belief Function Utility. From their Theorem 2.1, it follows that if DC is also assumed in the entry game example, then

$$\begin{array}{lcl} \frac{1}{2}B_1 + \frac{1}{2}N_2 & \succ_0 & \frac{1}{2}B_1 + \frac{1}{2}N_1 \iff \\ \frac{1}{2}B_1 + \frac{1}{2}B_2 & \succ_0 & B_1. \end{array}$$

That is, aversion to ambiguity about heterogeneity as in (2.6) is exhibited if and only if randomizing between the bets B_1 and B_2 has positive value.[17] But, as in the discussion of WOI, indifference to such randomization expresses a form of (conditional) stochastic independence across markets, which we take to be intuitive.[18]

Thus we would explain to the PM: "If your situation calls for 'sampling from some markets and then choosing a policy for the others', then you do not need full dynamic consistency and our model applies. Otherwise, you can't have everything. You will have to decide what you are willing to give up, and then we will have more modeling to do."

3. The Representation

3.1. Main result

We show in this section that the above axioms characterize the model outlined in the introduction, which we describe here in full generality.

In the preceding section, we defined belief functions on Ω via (2.2). In a similar fashion, replacing Ω and $\widehat{\Omega}$ by S and $\widehat{S}$, define belief functions on S, thought of as representing beliefs about a single experiment. Denote by $Bel\,(S)$ the set of all belief functions on S; a generic element is denoted θ.[19] Each θ induces a belief function on Ω, denoted θ^{∞} and referred to as the i.i.d. product of θ. Let $\left(\widehat{S}, \Gamma, m\right)$ generate θ as in (2.2) and consider the triple $(\left(\widehat{S}\right)^{\infty}, m^{\infty}, \Gamma^{\infty})$, where m^{∞} is the ordinary i.i.d. product of the probability measure m, and Γ^{∞} is the correspondence $\Gamma^{\infty} : \left(\widehat{S}\right)^{\infty} \rightsquigarrow \Omega = S^{\infty}$ given by

$$\Gamma^{\infty}\left(\widehat{s_1}, \widehat{s_2}, ...\right) = \Gamma\left(\widehat{s_1}\right) \times \Gamma\left(\widehat{s_2}\right) \times \; ... \tag{3.1}$$

[17] The cited theorem implies that $\succeq_0$ can be represented by a utility function that is additive across states, in particular, for any act depending only on $S_1 \times S_2$, $U_0\left(f\right) = \Sigma_{r \in S_1 \times S_2} v_r\left(f\left(r\right)\right)$. Then each of the indicated strict rankings is equivalent to the condition $v_{B_1 N_2}\left(\frac{1}{2}\right) + v_{N_1 B_2}\left(\frac{1}{2}\right) > v_{B_1 N_2}\left(1\right) + v_{N_1 B_2}\left(0\right)$.

[18] Indifference between $\frac{1}{2}B_1 + \frac{1}{2}B_2$ and B_1 follows in our model from Symmetry and (2.4), and the latter follows from Belief Function Utility and WOI.

[19] Endow $Bel\,(S)$ with the topology for which $\theta_n \to \theta$ if and only if $\int f d\theta_n \to \int f d\theta$ for every continuous function f on S, where the integral is in the sense of Choquet. Then $Bel\,(S)$ is compact metric.

Then θ^{∞} is the belief function on Ω generated as in (2.2) by $(\widehat{\Omega}, m^{\infty}, \Gamma^{\infty})$.[20] The construction (1.5) in the entry game example is a special case where $S = \{B, N\}$, $\theta = \theta_{\eta}$, $\widehat{S} = [0,1]^2$, m is the uniform measure on $[0,1]^2$ and Γ is the equilbrium correspondence.

The belief-function utility V on $\mathcal{F}$ is called an *i.i.d. (belief-function) utility* if there exists θ, a belief function on S, such that

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

When θ is additive, the function reduces to expected value with an i.i.d. probability measure.

Refer to L as a *likelihood function* if $L : Bel(S) \to \Delta(\Omega)$, where $\theta \longmapsto L(B \mid \theta)$ is (Borel) measurable for each measurable subset B of Ω.

The main result of the paper can now be stated.

Theorem 3.1. *Let $\{\succeq_{n,s^n} : n \geq 0,\ s^n \in S^n\}$ be preference orders on the set of acts $\mathcal{F}$. Then the statements (a) and (b) are equivalent:*

(a) $\{\succeq_{n,s^n}\}$ satisfies Belief Function Utility, Symmetry, Weak Orthogonal Independence, Consequentialism, Commutativity and Weak Dynamic Consistency.

*(b.i) **Choice:** For every n and s^n, there exists a (necessarily unique) Borel probability measure μ_{n,s^n} on $Bel(S)$ such that ν_{n,s^n}, the belief-function provided by Belief Function Utility, can be expressed in the form*

$$\nu_{n,s^n}(A) = \int_{Bel(S)} \theta^{\infty}(A)\, d\mu_{n,s^s}(\theta), \text{ for every Borel } A \subset \Omega, \tag{3.2}$$

and such that $\succeq_{n,s^n}$ is represented by U_{n,s^n} having the form

$$U_{n,s^n}(f) = \int_{Bel(S)} V_{\theta^{\infty}}(f)\, d\mu_{n,s^n}(\theta), \text{ for every } f \text{ in } \mathcal{F}. \tag{3.3}$$

*(b.ii) **Inference:** There exists a likelihood function $L : Bel(S) \to \Delta(\Omega)$ such that:*

[20] Appendix A shows that θ^{∞} is well defined (if $(\widehat{S}, m, \Gamma)$ and $(\widehat{S}', m', \Gamma')$ both generate θ, then they both lead to the same belief function on S^{∞}) and that it corresponds, in the case of finitely many experiments, to the product notion for belief functions proposed by Dempster (1967, 1968) and studied by Hendon et al. (1996).

L1 $\int_{Bel(S)} L\left(s^n \times \Omega \mid \theta\right) d\mu_0\left(\theta\right) > 0$ *for all* $n \geq 1$ *and* $s^n \in S^n$.

L2 Each $L\left(\cdot \mid \theta\right)$ *is exchangeable, that is, for every* $\theta \in Bel\left(S\right)$,

$$L\left(\cdot \mid \theta\right) = \int_{\Delta(S)} q^{\infty}\left(\cdot\right) d\lambda_{\theta}\left(q\right) \tag{3.4}$$

for some probability measure λ_{θ} *on* $\Delta\left(S\right)$.

L3 For each n *and* s^n, $\mu_n\left(\cdot \mid s^n\right)$ *is obtained by applying Bayes' rule to the prior* μ_0 *and the likelihood* L.

Focus on sufficiency of the axioms. Part (b.i) describes implications for preference conditional on any sample. That each conditional preference has a belief function on Ω, denoted ν_{n,s^n}, is an assumption through the axiom Belief Function Utility. The content of (b.i) is the structure of that belief function expressed in (3.2). The relation (1.6) described in the introduction is a special case. In the entry game example, each market is described by the parameter η, common to all markets, or by the corresponding belief function θ_{η} on $\{B, N\}$, which generates beliefs over Ω represented by the i.i.d. product $\left(\theta_{\eta}\right)^{\infty}$. Because η is unknown, a prior over η, and hence over θ_{η} is used and the expectation over all i.i.d. products describes beliefs over the sequence of markets. In the present general setting, markets are replaced by arbitrary experiments with outcomes in S, the unknown 'parameter' that is common across all experiments is θ, a belief function for each experiment, which is uncertain according to the prior or posterior μ_{n,s^n}, and the expectation over all i.i.d. products θ^{∞} represents beliefs over $\Omega = S^{\infty}$. De Finetti's celebrated 'conditionally i.i.d.' representation for exchangeable Bayesian beliefs is the special case of (3.2) where μ_{n,s^n} has support in the set of (additive) probability measures on $Bel\left(S\right)$.

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. In the Bayesian case, the parameter is the probability law that describes each experiment. Part (b.i) generalizes the notion of parameter to a belief function over each experiment, or equivalently, to its core, $core\left(\theta\right)$, a set of probability laws over a single experiment. In the Bayesian model, knowledge of the parameter determines a unique probability law over the sequence of experiments. In contrast, in our model knowledge of the parameter θ determines only a set, $core\left(\theta^{\infty}\right)$, of probability laws over the sequence of experiments. Because of the decision maker'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. In the entry game example, incompleteness is due to ignorance of how equilibria are selected. In general, incompleteness is revealed through preference and the underlying reasons are not modeled.

Thus far we have discussed part (b.i) and what it says about preference. It does not say how the posteriors μ_{n,s^n} are related to the prior μ_0, which is the inference component of the model and the content of part (b.ii). In the entry game example (Section 1.2), we indicated why the updating rule is not obvious and the same reason is apparent here: when updating beliefs about θ, there would seem to be many likelihood functions, one for each measure in $core\,(\theta)$, that could give the likelihood of observing a given experimental outcome conditional on θ. The model prescribes that one average over them in the sense of applying Bayes' rule to the likelihood function defined in (3.4). The measures λ_θ are subjective, as explained in Section 1.2, and unrestricted. Therefore, the complete model is defined by specifying both the prior μ_0 and the collection $\{\lambda_\theta : \theta \in Bel\,(S)\}$.

The fact that updating is Bayesian has the advantage that results from Bayesian learning theory translate directly. For example, in the entry game example, it follows from Acemoglu et al (2009) that for an exchangeable likelihood where each λ_η has support equal to $[0,\eta]$, then along some infinite samples posteriors can fail to converge to certainty about a single η, that is, η is only partially identified.[21]

Two final remarks in this section offer further perspective on Theorem 3.1.[22]

Remark 1. *The gap between our model and the exchangeable Bayesian model of inference and choice lies in the difference between Weak Dynamic Consistency and Dynamic Consistency. If we replace the former by the latter, then our model reduces to the exchangeable Bayesian model. (This follows from Epstein and Seo (2011, Thm. 2.1).) Note that if instead one strengthens WOI to the Independence axiom, then full dynamic consistency is still not implied and the two models differ in how they treat inference because only in our model is any exchangeable likelihood admissible.*

Remark 2. *A likelihood function of the form in (3.4) can be formulated for any abstract parameter θ. Therefore, one might wonder about the role of belief functions in the inference component of the model. However, the justification for (3.4)*

[21] See our paper (2010) for related illustrative results.

[22] The second is in response to a question raised by a referee.

provided by the theorem is based on the complete set of preference axioms, including Belief Function Utility, and is inextricably tied to the implication that belief functions are the correct parameters.

3.2. The entry game example again

As noted in Section 1.2 in the entry game example, given η and the uniform distribution on the unobservables ϵ_i, the set of logically possible probabilities for B equals the interval $[0, \eta]$. We modeled the PM's beliefs by assuming that she associates η with this interval, and hence with θ_η defined in (1.5). However, the model does not require this tight connection between subjective beliefs and what is logically possible.[23] In particular, according to the model the PM might even act as a Bayesian and as if assuming that, conditional on the value η, the probability that B is selected is $\eta/2$ and i.i.d. across markets.

More generally, beliefs are subjective and unobservable as is the analyst's underlying theory of selection. The definition of belief functions on S via triples $\left(\widehat{S}, m, \Gamma\right)$ permits a host of alternative theories of selection through alternative measures m, (m is after all also subjective), and through alternative auxiliary spaces and correspondences. For example, our model permits the probability interval corresponding to the parameter η to be a strict subset of $[0, \eta]$. Indeed, any interval $\left[\underline{p}, \overline{p}\right]$ defines a unique belief function: take the auxiliary state space $\widehat{S} = \{\{B\}, \{N\}, \{B, N\}\}$, the probability measure m given by $m(\{B\}) = \underline{p}$, $m(\{N\}) = 1 - \overline{p}$ and $m(\{B, N\}) = \overline{p} - \underline{p}$, and the correspondence Γ given by $\Gamma(\{B\}) = \{B\}$, $\Gamma(\{N\}) = \{N\}$ and $\Gamma(\{B, N\}) = \{B, N\}$. Then the counterpart of (2.2) defines the belief function θ for which

$$[\theta(B), 1 - \theta(N)] = \left[\underline{p}, \overline{p}\right].$$

With the generality of our model thus clarified, we can now verify that, apart from extreme special cases, the utility functions in (3.3) accommodate the rankings (2.5)-(2.7) that illustrate our intuition for the Jovanovic entry game. Suppose that μ_0 attaches positive probability to θ, where $\theta(B), \theta(N) > 0$ and $\theta(B) + \theta(N) < 1$. Then it suffices to show that the noted rankings are satisfied by the utility function V_{θ^∞}. This can be verified by straightforward calculations. For the third ranking,

[23] In the same way, the multiple-priors model of a decision maker confronting an Ellsberg urn does not require that her set of priors coincide with the set of all logically possible probability laws.

abbreviate the bet $\{B_1B_2, N_1N_2\}$ by f and let π be the permutation that switches the second the third markets. Then compute that

$$\begin{aligned} V_{\theta^\infty}(f) &= (\theta(B))^2 + (\theta(N))^2 \\ &< (\theta(B))^2 + (\theta(N))^2 + \theta(B)\,\theta(N)\,(1-\theta(B)-\theta(N)) \\ &= V_{\theta^\infty}\left(\tfrac{1}{2}f + \tfrac{1}{2}\pi f\right), \end{aligned}$$

which proves (2.7).

3.3. Prior beliefs and the LLN

Though Theorem 3.1 establishes the existence (given the axioms) of the subjective components μ_0 and $\{\lambda_\theta\}$ that pin down both preference and updating, it does not provide guidance to the decision maker as to how to arrive at these components. However, as we show here the model does provide a way for the decision maker 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. We illustrate this first in the entry game example and provide a general result in Appendix C. Admittedly, we address only μ_0, which is adequate if all choice is ex ante, and we leave open the question of how to provide guidance for arriving at the λ_θ's.

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.

Begin with a PM facing (1.1) who is certain that the selection probability of B 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 B converges to $q\eta$, and further that prior beliefs about η and the (certainty equivalent) utility for bets about empirical frequencies are related by[24]

$$\mu_0(\{\eta : 0 \leq \eta \leq \overline{\eta}\}) = U_0\left(\{\omega : \lim \Psi_n(\omega) \leq q\overline{\eta}\}\right). \tag{3.5}$$

(We consider the ex ante perspective only.) Therefore, the PM can calibrate her prior μ_0 over the parameter η if she can arrive at certainty equivalents for the indicated bets on limiting empirical frequencies.

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

Now consider a PM who is averse to her ignorance about selection as modeled in Section 1.2. 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 frequencies:[25]

$$\mu_0(\{\eta : 0 \leq \eta \leq \overline{\eta}\}) = U_0\left(\{\omega : \limsup \Psi_n(\omega) \leq \overline{\eta}\}\right). \tag{3.6}$$

In other words, the probability assigned to values of η no greater than $\overline{\eta}$ equals the certainty equivalent of the bet (with prizes 1 and 0) that, for all $\epsilon > 0$, the empirical frequency of B is less than $\overline{\eta} + \epsilon$ in all sufficiently large samples.[26]

See Appendix C for a more general result.

4. Prediction of empirical frequencies

We illustrate the choice component of our model and its tractability by applying it to an optimal point prediction problem, that of predicting optimally the empirical frequency of each outcome when the experiment has two possible outcomes. The entry game is one example; indeed we denote outcomes by B and N. The application serves also to illustrate the influence on decisions of unknown correlation and heterogeneity. In this binary case, each belief function θ can be identified with a probability interval $I_\theta = [\theta(B), \theta^*(B)]$ for B, where $\theta^*(B) = 1 - \theta(N)$. We permit $\theta(B) \neq 0$; recall Section 3.2.

Begin with beliefs μ about belief functions, (they may be prior beliefs or posteriors after observing a sample), and consider prediction for n markets. We model optimal prediction by the following decision problem:

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

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

[25] The proof (see Appendix C) is based on a LLN for i.i.d. products of belief functions due to Maccheroni and Marinacci (2005).

[26] Only the lim sup appears because we have assumed that η is associated with the probability interval $I_\eta = [0, \eta]$ having zero as its left endpoint. For other specifications of I_η, or equivalently, of θ_η, if $I_\eta \subset I_{\overline{\eta}} \iff \eta \leq \overline{\eta}$, then (3.6) is valid if the right side is replaced by $U_0\left(\{\omega : [\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset I_{\overline{\eta}}\}\right)$.

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

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

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 (C.1) and (C.2)). Fix θ and α and consider

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

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

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

for some P in $core(\theta^\infty)$. 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 $\theta(B)$ and $\theta^*(B)$, as in (4.2).

To gain some 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, consider the entry game and suppose that the only relevant belief functions are of the form θ_η satisfying

$$\theta_\eta(B) = 0 \text{ and } \theta_\eta^*(B) = \eta.$$

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:

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

and

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

One observation is that $\alpha_1 \neq \alpha_2 \neq \alpha_\infty$. Thus the optimal prediction depends on the number of markets being considered, which is intuitive when correlation between markets is a concern.

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

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

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

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

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^* satisfies on $\{B_1, N_1\} \times \{B_2, N_2\}$:[27]

$$P^*(B_i) = \eta,\ \ P^*(N_i) = 1 - \eta,\ \ i = 1, 2, \text{ for all } \eta, \tag{4.6}$$

and, if $\eta > \frac{1}{2}$,

$$P^*(B_1, B_2) = \eta,\ P^*(N_1, N_2) = 1 - \eta,\ P^*(B_1, N_2) = P^*(N_1, B_2) = 0. \tag{4.7}$$

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 B is selected in

[27] Only the first two markets matter and thus we consider only the marginal of P^* on $S_1 \times S_2$.

one market then it is certain to be selected also in the other, and similarly for N). Correlation does not play a role when predicting given $\eta < \frac{1}{2}$, where P^* is the i.i.d. product of the marginal in (4.6).

Another way to see the effect of correlation is by comparing our decision-maker with one who solves:

$$\max_{\alpha \in [0,1]} \inf_{m \in \Delta([0,\eta])} \int_{\Delta(S)} \int_{\Omega} - \left(\alpha - \Psi_n(\omega)\right)^2 dq^{\infty} dm(q). \tag{4.8}$$

This decision-maker is also uncertain about the probability in $[0,\eta]$ with which the outcome B is selected in any single market, but she differs in two respects from the one discussed above. She is certain that the selection mechanism is i.i.d. across markets, and for her the true selection probability (corresponding to q) is ambiguous - she cannot settle on a single distribution over $[0,\eta]$ and uses instead the set of all distributions on the interval.[28] Thus she resembles the decision-makers modeled in much of the robust Bayesian literature, and we refer to her as a robust Bayesian. When predicting the outcome in one market, the robust Bayesian makes the identical prediction (4.4) as our decision-maker. In fact, the two decision-makers would rank all bets on a single market identically because they have a common set of predictive priors on $\{B, N\}$, namely the set of all distributions for which the probability of B is no greater than η. However, they differ when predicting for two or more markets. In particular, $\alpha_2^{RB} \neq \alpha_2$.[29] We attribute this difference to the fact that only our agent is concerned about correlation between markets.[30]

Finally, it is interesting to note that the difference between predictions disappears when predicting for a very large number of markets. More precisely,

$$\lim_{n\to\infty} \alpha_n^{RB} = \lim_{n\to\infty} \alpha_n = \alpha_\infty = \frac{\eta}{2}.$$

Thus the effects of ambiguity about correlation vanish in the limit when predicting for a large number of markets.

[28] Identify $q \in \Delta(\{B, N\})$ with the point $q(B)$ in the unit interval.

[29] One can compute that $\alpha_2^{RB} = \eta$ if $\eta \leq \frac{1}{3}$ and $\frac{1+\eta}{4}$ otherwise.

[30] A concern with heterogeneity does not seem relevant: if one generalizes the objective function in (4.8) by allowing nonidentical product measures $q_1 \otimes q_2 \otimes ...$, the optimal prediction is not affected for any n.

5. Concluding Remarks

5.1. More general entry games

The model we have proposed can be applied to a large class of entry games. For example, consider the following payoff matrix where profits depend also on exogenous variables (player/market characteristics or policy variables), $x_i = (x_{i1}, x_{i2}) \in X \subset \mathbb{R}^{2K}$, assumed observable to both the players and the analyst. (Any finite number of players is easily accommodated.) The PM believes that, for each i, $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}) \in \mathcal{E} \subset \mathbb{R}^2$ is distributed according to m_α and that ϵ_i's are i.i.d. The full set of parameters is $\phi = (\alpha, \beta_1, \beta_2, \eta) \in \Phi$.

	out	in
out	$0, 0$	$0, \beta_2 x_{i2} + \epsilon_{i2}$
in	$\beta_1 x_{i1} + \epsilon_{i1}, 0$	$\beta_1 x_{i1} + \eta + \epsilon_{i1}, \beta_2 x_{i2} + \eta + \epsilon_{i2}$

Let $Y = \{\text{out,in}\}^2$ be the set of all pure strategy profiles in any single market. Given ϕ, the basic uncertainty concerns which pure strategy Nash equilibrium will be played for each given x_i. Thus describe the set of outcomes for each market by

$$S = Y^X,$$

and denote the equilibrium correspondence by $\Gamma_\phi : \mathcal{E} \rightsquigarrow S$. Then the triple $(\mathcal{E}, m_\alpha, \Gamma_\phi)$ defines a belief function θ_ϕ on S.

The rest of the specification proceeds as before with one modification for inference. The state for market i would not be revealed by any real data. Rather, one would observe for market i the equilibrium y_i and the associated value x_i. Not observing entry decisions for other values x_i' can be captured by modeling the signal forthcoming from market i by the event

$$\left\{s \in Y^X : s(x_i) = y_i\right\} \subset S_i.$$

Then updating can be modeled as above but using the modified filtration on $S_1 \times S_2 \times ...$ that is defined thereby.

The more general class of games permits policy tools that affect profits. Moreover, the choice between such policy tools translates into a choice between acts and thus policy decisions can be modeled using belief function utility. As an example, suppose that the PM can choose between the policy variables x_1^* and x_1^{**} for market 1. They correspond to the acts f^*, f^{**} defined on the full state space

$\Pi_{i=1}^{\infty} Y^X$, where

$$f^*(s_1, ..., s_i, ...) = u(s_1(x^*), x^*) \text{ and } f^{**}(s_1, ..., s_i, ...) = u(s_1(x^{**}), x^{**}).$$

Here $s_1(x^*)$ is the Nash equilibrium strategy profile in market 1 given state $s_1 \in Y^X$, and $u(\cdot)$ gives the payoff to the policy maker as a function of the Nash equilibrium profile and the policy variable. A similar interpretation applies to $s_1(x^{**})$.

5.2. More related literature

Epstein and Seo (2010, Thm. 5.2) extend the de Finetti theorem to the class of multiple-priors preferences. Belief function utility is appealing because it is a special case of both multiple-priors utility and Choquet expected utility, and thus is "close" to the benchmark expected utility model. This closeness permits a much sharper representation result here in permitting both much simpler axioms and a stronger representation. The latter point concerns the meaning of "stochastic independence." Stochastic independence is more complicated in the nonadditive probability (or multiple-priors) framework and there is more than one way to form independent products (Ghirardato (1997)). Accordingly, the representation in our previous paper admits various ways of forming i.i.d. products. In contrast, in our model the rule for forming the i.i.d. product θ^{∞} is pinned down - it corresponds to that advocated by Dempster (1967, 1968) and Hendon et al. (1996). To our knowledge, this paper is the first to provide (via Theorem 3.1) a choice-theoretic rationale for any particular i.i.d. product rule.[31] The value added herein lies also in the demonstrated usefulness of employing ambiguity averse preferences to accommodate issues arising from theory incompleteness.

There exist a number of other generalizations of the de Finetti theorem to ambiguity averse preferences; see Epstein and Seo (2010, Thm. 3.2), Al Najjar and de Castro (2010), Cerreia-Vioglio et al. (2011) and Klibanoff et al. (2011). They are all in the spirit of what we referred to as the robust Bayesian model (recall (4.8)), in that they deal with ambiguity about parameters but exclude ambiguity about how experiments are related; for example, they cannot exhibit the rankings (2.6) and (2.7). As a result these models seem orthogonal to the central issues

[31] Ghirardato (1997) shows that the Hendon rule is the only product rule for belief functions such that the product (i) is also a belief function, and (ii) it satisfies a mathematical property called the Fubini property. In our model, this property emerges as an implication of assumptions about preference.

raised by multiple equilibria in entry games. Moreover, they address choice but not updating. A model in the same spirit is found in Shafer (1982), who is the first, to our knowledge, to discuss the use of belief functions within the framework of parametric statistical models analogous to de Finetti's. He sketches (section 3.3) a de Finetti-style treatment of randomness based on belief functions. His model is not axiomatic or choice-based.

When experiments are ordered in time, Epstein and Schneider (2007, 2008), model learning and choice under ambiguity using a specification for utility inspired by de Finetti's. They posit functional forms without foundations and motivate them through applications. Their model violates Symmetry and thus is not suited for cross-sectional applications such as discussed here.

5.3. Statistical decision theory

Finally, we relate decision problems as modeled here to those considered in statistical decision theory.[32] Conditional decisions in our model are determined by optimization problems of the form

$$\max_{f\in\Upsilon}\int_{Bel(S)}\left[\int_{\Omega}fd\theta^{\infty}\right]d\mu_{N}\left(\theta\mid s^{N}\right), \tag{5.1}$$

where Υ is the feasible set of acts and $\mu_N\left(\cdot\mid s^N\right)$ is the posterior after having observed the outcomes $s^N=(s_1,...,s_N)$ of the first N experiments. Fix $N\geq 0$; if $N=0$, then choice is made ex ante without the benefit of a sample. Without loss of generality, the payoffs to acts in Υ depend only on the outcomes of the remaining experiments $i=N+1,N+2,...$

Such conditional choices can be readily translated into the formalism of statistical decision theory. Define a (feasible) *decision rule* δ as a mapping $\delta:S^N\rightarrow\Upsilon$. The set of all such decision rules is $\mathbb{D}=\Upsilon^{S^N}$. Then the collection of problems (5.1), with s^N varying over all possible samples, can be reformulated as one of optimization over decision rules:

$$\max_{\delta\in\mathbb{D}}\int_{Bel(S)}u\left(\delta,\theta\right)d\mu_{0}\left(\theta\right)=-\min_{\delta\in\mathbb{D}}\int_{Bel(S)}r\left(\delta,\theta\right)d\mu_{0}\left(\theta\right), \tag{5.2}$$

where

$$-r\left(\delta,\theta\right)=u\left(\delta,\theta\right)=\Sigma_{s^N}L\left(s^N\mid\theta\right)\left(\int_{\Omega}\delta\left(s^N\right)\left(\cdot\right)d\theta^{\infty}\left(\cdot\right)\right). \tag{5.3}$$

[32]The latter are formulated below following Stoye (2012) in order to facilitate comparison with his approach to exploiting ambiguity modeling for statistical decision-making.

(Here L is any likelihood used for updating as provided by Theorem 3.1.) The function $r(\delta, \cdot)$ is the *risk function* of decision rule δ. Its expectation using μ_0 is the *Bayes risk* of δ. Thus conditional optimization in our model translates into the common prescription that an optimal decision rule should minimize Bayes risk.

This prescription seems entirely standard and, in particular, might appear to exclude any role for ambiguity. However, from this perspective, the novelty in our model is two-fold. First, in statistical decision theory θ is taken to lie in an abstract set Θ of parameters. The set Θ is a primitive, it presumably comes with the statistical decision-maker, and no justification is attempted for adopting any particular parametrization. In contrast, for our setting we have justified taking the parameter set to be the set of belief functions on S (the set of outcomes for a single experiment); recall that when S is binary, this is equivalent to taking intervals within $[0, 1]$ as the relevant parameters. The second novelty is that we justify the particular form (5.3) for the risk function. For example, in the prediction problem of Section 4, in the "no-data" case $N = 0$, a decision rule δ is simply a point forecast, $\mathbb{D} = [0, 1]$, and the risk function takes the form

$$r(\delta, \theta) = -\int_{\Omega} G(\Psi_n(\omega) - \delta)\, d\theta^{\infty}.$$

It is through the specific form for r, notably through the i.i.d. products of belief functions and Choquet expectation, that aversion to unknown correlation and heterogeneity is captured. Both achievements are possible because we take as primitives preferences over bets (or acts) whose payoffs depend on the realized outcomes of experiments rather than on the true value of a parameter. We emphasize that real world decision problems under uncertainty are generally of this form. For example, investment choice is a bet that realized returns will be favorable rather than a bet on the true mean and variance of the underlying distribution of returns; and the payoff to a policy choice in the entry game example depends on the entry outcomes realized in the relevant markets and only indirectly on the value of η.

In contrast, the axiomatic modeling reviewed in Stoye (2012) takes the risk function r as a primitive and hence unexplained. As a result, it cannot address the issues addressed here - ambiguity about correlation and heterogeneity arising from the multiplicity of equilibria in entry games. In fact, the formulation described by Stoye does not relate specifically to the setting of repeated experiments. Rather it deals in an abstract setting with how uncertainty about risk measures (due to uncertainty about θ) is, or should be, evaluated by the statistical analyst. For example, in addition to the Bayesian criterion as in (5.2), Stoye describes axiomatizations of Γ-minimax loss (where the single prior μ_0 is replaced by a set of priors as in the Gilboa-Schmeidler model), and versions of minimax regret.

A. Appendix: 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 probability measures on Ω endowed with the weak convergence topology; and $Bel(\Omega)$ is the space of belief functions endowed with the topology for which $\nu_n \to \nu$ if and only if $\int f d\nu_n \to \int f d\nu$ for every continuous function f on 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 $\nu : \Sigma_\Omega \to [0,1]$ satisfying:

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

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

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

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

Bel.5 ν is totally monotone (or ∞-monotone): for all Borel sets $B_1, .., B_n$,

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

These conditions are adapted from Phillipe et al. (1999). Conditions Bel.1-Bel.4 form a common definition of capacity (Schmeidler (1989)). 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 ν is convex (supermodular, or 2-alternating).

An important result regarding belief functions is Choquet's Theorem. Our statement of the theorem relies on Phillipe et al. (1999, Thms. 2 and 3), Molchanov (2005, Thm. 5.1) and Castaldo et al. (2004, Thm. 3.2). Note that, by Phillipe at al. (1999, 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.[33]

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

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

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

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

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 (2.2). (We note that the latter formulation is due to Dempster (1967) and Shafer (1976).) For one direction, Bel.1-Bel.5 imply the representation (A.1), which is the special case of (2.2) where $\widehat{\Omega} = \mathcal{K}(\Omega)$, Γ maps any K (a point in $\mathcal{K}(\Omega)$) into K (a subset of Ω) and $m = m_\nu$. Conversely, let ν be defined via the triple $\left(\widehat{\Omega}, m, \Gamma\right)$ and (2.1)-(A.4). View Γ as a function from $\widehat{\Omega}$ to $\mathcal{K}(\Omega)$. Then Γ is measurable (Aliprantis and Border (2006, Thm. 18.10)) and induces the measure $m' = m \circ \Gamma^{-1}$ on $\mathcal{K}(\Omega)$. Then Choquet's Theorem implies that $\nu(\cdot) = m \circ \Gamma^{-1}(\{K : K \subset \cdot\})$ satisfies Bel.1-Bel.5 and $m' = m_\nu$.

[33] 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 $\int_X f dP$.

Associated with any belief function ν is its core defined by

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

Then[34]

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

Turn to the corresponding utility function. The objects of choice are (Borel measurable) acts $f : \Omega \to [0,1]$, which for simplicity are restricted to have finite range (such acts are commonly called 'simple'). The utility $U(f)$ of any act f is defined by (2.3). By Molchanov (2005, 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 multiple-priors model (Gilboa and Schmeidler (1989)) with set of priors equal to $core(\nu)$:

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

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 (1989) refer to these properties as certainty additivity and ambiguity aversion respectively. We use them repeatedly.

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

[34] When the support of m is not finite, a measurability assumption for $\widehat{\omega} \longmapsto p_{\widehat{\omega}}$ must be added to give meaning to this expression.

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

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

Now let $\Omega = S^\infty$. Let $\theta \in Bel\,(S)$ be generated by $(\widehat{S}, m, \Gamma)$. We defined θ^∞ to be the belief function on Ω represented by $(\widehat{\Omega}, m^\infty, \Gamma^\infty)$, where: $\widehat{\Omega} = \left(\widehat{S}\right)^\infty$, m^∞ is the ordinary i.i.d. product of the probability measure m, and Γ^∞ is the correspondence $\Gamma^\infty : \widehat{\Omega} \rightsquigarrow \Omega = S^\infty$ given by (3.1). Choquet's theorem gives an alternative characterization of the product that we use frequently. In particular, it implies that the product θ^∞ does not depend on the particular representation $(\widehat{S}, m, \Gamma)$ for θ.

Lemma A.2. *Let $\theta \in Bel\,(S)$ correspond to $m_\theta \in \Delta\,(\mathcal{K}\,(S))$ as in Choquet's theorem. Then $\theta^\infty \in Bel\,(\Omega)$ is the unique belief function corresponding to $(m_\theta)^\infty \in \Delta\,(\mathcal{K}\,(\Omega))$ as in Choquet's theorem (where $(m_\theta)^\infty$ is the i.i.d. product of the measure $(m_\theta)^\infty$).*

The proof of the lemma is omitted. Note that $(m_\theta)^\infty$ 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 $\mathcal{K}\,(\Omega)$.

By Philippe et al. (1999, Thm. 3),

$$core\,(\theta^\infty) = \int \Delta\,(\Gamma\,(\epsilon_1) \times \Gamma\,(\epsilon_2) \times \ ...)\, dm^\infty\,(\epsilon_1, ...) \tag{A.8}$$

where the integral is an Aumann integral. This characterization of the core was used to derive the implication (1.8).

B. Appendix: Proof of Theorem 3.1

First we prove the measurability required to show that the integrals in (3.2) and (3.3) 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. *Both $\theta \longmapsto V_{\theta^\infty}\,(f)$ and $\theta \longmapsto \theta^\infty\,(A)$ are universally measurable for any $f \in \mathcal{F}$ and $A \in \Sigma_\Omega$.*

Proof. Since $Bel\left(S\right)$ and $\Delta\left(\mathcal{K}\left(S\right)\right)$ are homeomorphic, and in light of (A.2), it is enough to prove analytical (and hence universal) measurability of the mapping from $\Delta\left(\mathcal{K}\left(S\right)\right)$ to $\mathbb{R}$ given by

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

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

Step 2. $P \longmapsto \int \hat{f}dP$ from $\Delta\left([\mathcal{K}\left(S\right)]^\infty\right)$ to $\mathbb{R}$ is analytically measurable for any bounded analytically measurable function $\hat{f}$ on $[\mathcal{K}\left(S\right)]^\infty$: If $\hat{f}$ is simple (has a finite number of values), then $P \longmapsto \int \hat{f}dP$ is analytically measurable by Aliprantis and Border (2006, p. 169). More generally, $\int \hat{f}dP$ equals the pointwise limit of $\lim \int \hat{f}_n dP$ for some simple and analytically measurable $\hat{f}_n$, which implies the desired measurability.

Step 3. Note that

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

is coanalytic by Phillipe et al. (1999, p. 772), and hence analytically measurable.

Steps 1, 2 and 3 complete the proof. ■

Where conditioning is not important, it is suppressed notationally and we refer to a generic ν, U and μ. For any $\nu \in Bel\left(\Omega\right)$, denote by $\zeta\left(\nu\right)$ the measure m_ν on $\mathcal{K}\left(\Omega\right)$ provided by the Choquet theorem. (Similarly if $\theta \in Bel\left(S\right)$, then $\zeta\left(\theta\right) \in \Delta\left(\mathcal{K}\left(S\right)\right)$.) We use (A.2) repeatedly without reference.

Necessity of the axioms. $(b) \implies (a)$: Belief Function Utility is obvious. Verify that V_{θ^∞} satisfies Symmetry and WOI, which implies the same for U. Let $m = \zeta\left(\theta^\infty\right)$. By Lemma A.2, m is an i.i.d. measure on $[\mathcal{K}\left(S\right)]^\infty$, hence symmetric. Therefore,

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

Show (2.4) to prove WOI. For simplicity, let $f \in \mathcal{F}_1$ and $g \in \mathcal{F}_2$. The general case is similar. For $0 < \alpha \leq 1$,

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

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

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

Necessity of Consequentialism, Commutativity and WDC is readily verified.

Sufficiency of the axioms. $(a) \Longrightarrow (b.i)$: Show that axioms $\Longrightarrow$ (3.2) $\Longrightarrow$ (3.3).

Proof that (3.2) $\Longrightarrow$ (3.3): Let Σ' be the σ-algebra generated by the class

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

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

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

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

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

which is true given (3.2).

By a standard argument using the Lebesgue Dominated Convergence Theorem,

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

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

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

Proof that axioms $\Longrightarrow$ (3.2): 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.2. *For any $m \in \Delta(\mathcal{K}(\Omega))$, $m = \pi m$ for all π if and only if $m = \zeta(\nu)$ for some symmetric belief function ν on Ω.*

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

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

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 $\mathcal{K}(\Omega)$. ■

Lemma B.3. *Let ν be a belief function on Ω and $m = \zeta(\nu)$ the corresponding measure on $\mathcal{K}(\Omega)$. If U_ν satisfies WOI, then $m\left[(\mathcal{K}(S))^\infty\right] = 1$.*

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 $\{1, ..., n\} = I \cup J$,

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

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

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

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

$a.s.$-$m\,[K]$: This is where WOI enters - by (2.4) it implies that

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

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

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

The assertion follows from

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

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

Step 2. $m\left(\bigcap_{(f,g) \in \mathcal{G}} \mathcal{B}_{f,g}\right) = 1$: Since the set of continuous finitely-based acts is separable under the sup-norm topology (Aliprantis and Border (2006, Lemma 3.99)), it is easy to see that $\mathcal{G}$ is also separable. Let $\{(f_n, g_n)\}$ be a countable dense subset of $\mathcal{G}$. By Step 1,

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

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

Only $\subset$ requires proof. Let $K \in \bigcap_{i=1}^{\infty} \mathcal{B}_{f_i, g_i}$, $(f, g) \in \mathcal{G}$ and assume without loss of generality that $(f_i, g_i) \to (f, g)$. Then, by the Maximum Theorem (Aliprantis and Border (2006, Thm. 17.31),

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since $K \in \mathcal{A}$ and $\omega^1, \omega^2 \in K$, there exists $\hat{\omega}^2 \in K$ such that $\left(\hat{\omega}_1^2, \hat{\omega}_2^2\right) = (\omega_1^1, \omega_2^2)$. Similarly, since $\hat{\omega}^2, \omega^3 \in K$, there exists $\hat{\omega}^3 \in K$ such that $\left(\hat{\omega}_1^3, \hat{\omega}_2^3, \hat{\omega}_3^3\right) = \left(\hat{\omega}_1^2, \hat{\omega}_2^2, \omega_3^3\right) = (\omega_1^1, \omega_2^2, \omega_3^3)$, and so on, giving a sequence $\{\hat{\omega}^n\}$ in K. Any limit point ω^* satisfies (B.4). ■

Let ν be a belief function on Ω and suppose that U_ν satisfies Symmetry and WOI. By Lemma B.3, $m \equiv \zeta(\nu)$ can be viewed as a measure on $[\mathcal{K}(S)]^{\infty}$, and by Lemma B.2, m is symmetric. Thus we can apply de Finetti's Theorem (Hewitt and Savage (1955)) to m, viewing $\mathcal{K}(S)$ as the one-period state space, to obtain: There exists $\hat{\mu} \in \Delta(\Delta(\mathcal{K}(S)))$ such that

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

Here each 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

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

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

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

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

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

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

$(a) \Longrightarrow (b.ii)$: Existence of a likelihood function $L : Bel(S) \rightarrow \Delta(S^{\infty})$ satisfyng $L1$ and $L3$ amounts to a minor modification of Epstein and Seo (2010, Thm. 6.1) and can be proven similarly. (The latter adopts multiple-priors utility as the framework, rather than belief function utility. However, this difference is of no significance for the proof and calls only for an obvious translation.)

It remains only to prove that $L2$ (exchangeability for L) follows from Commutativity. While L need not be exchangeable, we prove that there exists an exchangeable likelihood L^{*} that also generates updating.

Assume $S = \{B, N\}$ for notational simplicity. By Commutativity, $U_n(f \mid s^n) = U_n(f \mid \pi s^n)$ for all f, n, s^n and π. By uniqueness of the representing measure noted above,

$$\mu(\cdot \mid s^n) = \mu(\cdot \mid \pi s^n).$$

Let

$$\overline{L}(s^n) \equiv \int L(s^n \mid \theta)\, d\mu_0.$$

Then

$$\begin{aligned}
U_0(f) &= \sum_{s^n} \overline{L}(s^n) U_n(f \mid s^n) \\
&= \sum_{k=0}^{n} \sum_{s^n \in S^{n,k}} \overline{L}(s^n) U_n(f \mid s^n) \\
&= \sum_{k=0, S^{n,k} \neq \varnothing}^{n} \left(\sum_{s^n \in S^{n,k}} \overline{L}(s^n) \right) U_n(f \mid s^n),
\end{aligned}$$

where $S^{n,k}$ is the set of all samples $s^n \in S^n$ with k occurrences of B. Define $\overline{L}_n^* \in \Delta(S^n)$ by

$$\overline{L}_n^*(s^n) = \frac{1}{|S^{n,k}|} \left(\sum_{s^n \in S^{n,k}} \overline{L}(s^n) \right) \text{ if } s^n \in S^{n,k}.$$

By the Kolomogorov Extension Theorem, there exists $\overline{L}^* \in \Delta(S^\infty)$ that coincides with $\overline{L}_n^*$ on S^n for every n. Therefore, it is exchangeable and satisfies, for every n,

$$U_0(\cdot) = \sum_{s^n} \overline{L}^*(s^n) U_n(\cdot \mid s^n) \text{ on } \mathcal{F}. \tag{B.5}$$

The latter equation leads to the desired likelihood function. Take

$$L^*(s^n \mid \theta) \equiv \overline{L}^*(s^n) \left(d\mu(\theta \mid s^n) / d\mu_0(\theta) \right).$$

By the uniqueness of representing measures in Theorem 3.1, (B.5) implies

$$\mu_0(\cdot) = \sum_{s^n} \overline{L}^*(s^n) \mu(\cdot \mid s^n),$$

and thus $\Sigma_{s^n} L^*(s^n \mid \theta) = 1$ for all θ. Further,

$$\begin{aligned}
L^*(s^n \mid \theta) &= \overline{L}^*(s^n) \left(d\mu(\theta \mid s^n) / d\mu_0(\theta) \right) \\
&= \overline{L}^*(\pi s^n) \left(d\mu(\theta \mid \pi s^n) / d\mu_0(\theta) \right) \\
&= L^*(\pi s^n \mid \theta).
\end{aligned}$$

It is easily verified that posteriors are generated by Bayesian updating using μ_0 and L^* (proceed as in the proof of Theorem 6.1 of our earlier paper). ■

C. Appendix: Proofs for Section 3.3

Proof of (3.6): The LLN in Maccheroni and Marinacci (2005) implies that

$$(\theta_\eta)^\infty \left(\{\omega \in \Omega : [\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset I_\eta\}\right) = 1. \tag{C.1}$$

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

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

Therefore, the representation (3.2) implies that, for every $0 \leq b \leq 1$,

$$\begin{aligned} &\mu_0\left(\{\eta : I_\eta \subset [0,b]\}\right) \\ = \ &U_0\left(\{\omega : [\liminf \Psi_n(\omega), \limsup \Psi_n(\omega)] \subset [0,b]\}\right) \\ = \ &U_0\left(\{\omega : \limsup \Psi_n(\omega) \leq b\}\right). \end{aligned} \tag{C.3}$$

Take $[0,b] = I_{\overline{\eta}}$. Because $I_\eta \subset I_{\overline{\eta}}$ if and only if $\eta \leq \overline{\eta}$, (3.6) follows. ■

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

Corollary C.1. *Adopt the assumptions in Theorem 3.1 and let U_0 and μ_0 be as provided there. Then:*
(a) For every finite collection $\{A_1, ..., A_J\}$ of subsets of S, and for all $a_j \leq b_j$, $j = 1, ..., J$,

$$\begin{aligned} &\mu_0\left(\bigcap_{j=1}^{J} \{\theta : [\theta(A_j), 1 - \theta(S \backslash A_j)] \subset [a_j, b_j]\}\right) \\ = \ &U_0\left(\bigcap_{j=1}^{J} \{\omega : [\liminf \Psi_n(A_j)(\omega), \limsup \Psi_n(A_j)(\omega)] \subset [a_j, b_j]\}\right). \end{aligned} \tag{C.4}$$

(b) Let μ' be any probability measure on $Bel(S)$ that agees with μ_0 on all sets of the form

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

where A_j, a_j and J vary over the nonempty subsets of S, $[0,1]$ and the positive integers respectively. Then $\mu' = \mu_0$.

Equation (C.4) relates the prior μ_0 over parameters, here belief functions, to the evaluation of bets on empirical frequencies for the events $A_1, ..., A_J$. More precisely, the μ_0-measures of the sets shown are so related. However, (b) shows that μ_0 is completely determined by its values on these sets.

We need 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 C.2. *Let $K \in [\mathcal{K}(S)]^\infty$, $K = K_1 \times K_2 \times ...$, and $\alpha \in \mathbb{R}$. Then the following are equivalent:*
(i) $\liminf_n \Psi_n(A)(\omega) > \alpha$ *for every* $s_i \in K_i$, $i = 1, ...$
(ii) $\liminf_n \widehat{\Psi}_n(A)(K) > \alpha$.

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

(ii)⇒(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 $i = 1, 2, ...$, then,

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

■

Lemma C.3. *(i)* $\theta^\infty(\{\omega : \theta(A) < \liminf_n \Psi_n(A)(\omega)\}) = 0$ *for each* $A \subset S$*; and (ii)* $\theta^\infty(\{\omega : \limsup_n \Psi_n(A)(\omega) < 1 - \theta(S\backslash A)\}) = 0$ *for each* $A \subset S$.

Proof. Fix $A \subset S$. Then,

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

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

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

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

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

By the LLN in Maccheroni and Marinacci (2005), Lemma C.3 and the monotonicity of belief functions,

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

and

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

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,

$$\begin{aligned} & \nu\left(\bigcap_{j=1}^{J}\left\{\omega : \left[\liminf \Psi_n\left(A_j\right)\left(\omega\right), \limsup \Psi_n\left(A_j\right)\left(\omega\right)\right] \subset [a_j,b_j]\right\}\right) \\ = & \int_{Bel(S)} \theta^{\infty}\left(\bigcap_{j=1}^{J}\left\{\omega : \left[\liminf \Psi_n\left(A_j\right)\left(\omega\right), \limsup \Psi_n\left(A_j\right)\left(\omega\right)\right] \subset [a_j,b_j]\right\}\right) d\mu_0\left(\theta\right) \\ = & \mu_0\left(\bigcap_{j=1}^{J}\left\{\theta : \left[\theta\left(A_j\right), 1-\theta\left(S \backslash A_j\right)\right] \subset [a_j,b_j]\right\}\right). \end{aligned}$$

(b) We can identify μ' and μ_0 with measures on $\Delta(\mathcal{K}(S))$. Modulo this identification, we are given that μ' and μ_0 agree on the collection of all subsets of $\Delta(\mathcal{K}(S))$ of the form

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

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. $\ell \longmapsto \ell(C)$ 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 $\ell \longmapsto \ell(C)$ 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 $\mathcal{K}(S) \backslash C \in \mathcal{C}$; (ii) if each $C_j \in \mathcal{C}$, then $\ell \longmapsto \ell\left(\cup_{j=1}^{\infty} C_j\right)$ is Σ^*-measurable because it equals the pointwise limit of $\ell \longmapsto \ell\left(\cup_{j=1}^{n} C_j\right)$ - hence $\cup_{j=1}^{\infty} C_j \in \mathcal{C}$.

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

D. Appendix: Prediction

This appendix deals with binary experiments, $S = \{B, N\}$. Each belief function θ on S corresponds to the probability interval for outcome B given by $[\theta(B), \theta^*(B)]$. The entry game is one example but here we do not impose $\theta(B) = 0$. The empirical frequency of B in the first n experiments of the sample $\omega \in S^\infty$ is denoted $\Psi_n(\omega)$.

We make use of the following Central Limit Theorem (CLT) for belief functions (Epstein and Seo (2011b)).

Theorem D.1 (CLT). *Suppose that $G_n : \mathbb{R} \to \mathbb{R}$ is quasiconcave and continuous for each n and that $\sup_{n,t} |G_n(t)| < \infty$. Let (X_{1n}, X_{2n}) be normally distributed with mean $(\theta(B), \theta^*(B))$ and variance*

$$\frac{1}{n}\begin{pmatrix} \theta(B)(1-\theta(B)) & \theta(B)\theta(N) \\ \theta(B)\theta(N) & (1-\theta(N))\theta(N) \end{pmatrix}.$$

Then

$$\int G_n(\Psi_n(\omega))\, d\theta^\infty = E\left[\min\left\{G_n(X_{1n}), G_n(X_{2n})\right\}\right] + O\left(\frac{1}{\sqrt{n}}\right),$$

that is, there exists a constant K such that

$$\limsup_{n\to\infty} \sqrt{n}\left|\int G_n(\Psi_n(\omega))d\theta^\infty - E\left[\min\left\{G_n(X_{1n}), G_n(X_{2n})\right\}\right]\right| \leq K.$$

Proof of Theorem 4.1: Step 1: Show that

$$\begin{aligned}
&\lim_{n\to\infty}\int\int G\left(\alpha-\Psi_n\left(\omega\right)\right)d\theta^{\infty}d\mu_0\left(\theta\right)\\
=\;&\int\min\left\{G\left(\alpha-\theta\left(B\right)\right),G\left(\alpha-\theta^*\left(B\right)\right)\right\}d\mu_0\left(\theta\right).
\end{aligned}$$

By the CLT, for each θ,

$$\int G\left(\alpha-\Psi_n\left(\omega\right)\right)d\theta^{\infty}=E\left[\min\left\{G\left(\alpha-\bar{X}_{1n}\right),G\left(\alpha-\bar{X}_{2n}\right)\right\}\right]+O\left(\frac{1}{\sqrt{n}}\right),$$

where $\bar{X}_{1n}=\frac{1}{n}\sum_{i=1}^{n}X_{1i}$ and $\bar{X}_{2n}=\frac{1}{n}\sum_{i=1}^{n}X_{2i}$, and each $\left(X_{1i},X_{2i}\right)$ is normally distributed (i.i.d. across i's) with mean $\left(\theta\left(H\right),1-\theta\left(T\right)\right)$ and variance

$$\begin{pmatrix}\theta\left(B\right)\left(1-\theta\left(B\right)\right) & \theta\left(B\right)\theta\left(N\right)\\ \theta\left(B\right)\theta\left(N\right) & \left(1-\theta\left(N\right)\right)\theta\left(N\right)\end{pmatrix}.$$

By the classical strong LLN, $\left(\bar{X}_{1n},\bar{X}_{2n}\right)$ converges to $\left(\theta\left(B\right),\theta^*\left(B\right)\right)$ a.s. with respect to the i.i.d. product of the above normal. Then, by the continuous mapping theorem, $\min\left\{G\left(\alpha-\bar{X}_{1n}\right),G\left(\alpha-\bar{X}_{2n}\right)\right\}$ converges to $\min\left\{G\left(\alpha-\theta\left(B\right)\right),G\left(\alpha-\theta^*\left(B\right)\right)\right\}$ a.s. and thus in distribution. Therefore,

$$E\left[\min\left\{G\left(\alpha-\bar{X}_{1n}\right),G\left(\alpha-\bar{X}_{2n}\right)\right\}\right]\to\min\left\{G\left(\alpha-\theta\left(B\right)\right),G\left(\alpha-\theta^*\left(B\right)\right)\right\},\tag{D.1}$$

and $\int G\left(\alpha-\Psi_n\left(\omega\right)\right)d\theta^{\infty}\to\min\left\{G\left(\alpha-\theta\left(B\right)\right),G\left(\alpha-\theta^*\left(B\right)\right)\right\}$. Apply the Dominated Convergence Theorem to complete the proof.

Step 2: Show that

$$\begin{aligned}
&\lim_{n\to\infty}\operatorname*{argmax}_{\alpha\in[0,1]}\int\int G\left(\alpha-\Psi_n\left(\omega\right)\right)d\theta^{\infty}d\mu\left(\theta\right)\\
=\;&\operatorname*{argmax}_{\alpha\in[0,1]}\lim_{n\to\infty}\int\int G\left(\alpha-\Psi_n\left(\omega\right)\right)d\theta^{\infty}d\mu\left(\theta\right).
\end{aligned}$$

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

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

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

F is well-defined by Step 1. It is also jointly continuous: We need to check only the case $\alpha_n \to \alpha$ and $n \to \infty$. 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,

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

References

[1] D. Acemoglu, V. Chernozhukov and M. Yildiz, Fragility of asymptotic agreement under Bayesian learning, 2009.

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

[3] N. Al Najjar and L. de Castro, Parametric representation of preferences, 2010.

[4] F.J. Anscombe and R.J. Aumann, A definition of subjective probability, *Ann. Math. Stats.* 34 (1963), 199-205.

[5] P. Bajari, H. Hong, and S.P. Ryan, Identification and estimation of a discrete game of complete information, *Econometrica* (2010) 78, 1529-1568.

[6] A. Beresteanu, I. Molchanov and F. Molinari, Sharp identification regions in models with convex moment predictions, *Econometrica* (2011) 79, 1785-1821.

[7] A. Beresteanu and F. Molinari, Asymptotic properties for a class of partially identified models, *Econometrica* (2008) 76, 763-784.

[8] S. Berry, Estimation of a model of entry in the airline industry, *Econometrica* (1992) 60, 889-917.

[9] S. Berry and E. Tamer, Identification in models of oligopoly entry, *Advances in Economics and Econometrics. Ninth World Congress of the Econometric Society* (2006) 2, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press, 46-85.

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

[11] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Ambiguity and robust statistics, 2011.

[12] V. Chernozhukov, H. Hong and E. Tamer, Estimation and confidence regions for parameter sets in econometric models, *Econometrica* 75 (2007), 1243-1284.

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

[14] F. Ciliberto and E. Tamer, Market structure and multiple equilibria in airline markets, *Econometrica* 77 (2009), 1791-1828.

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

[16] A. P. Dempster, A generalization of Bayesian inference, *J. Royal Stat. Soc.* Series B (1968) 30, 205-232.

[17] F. Echenique and I. Komunjer, Testing models with multiple equilibria by quantile methods, *Econometrica* 77 (2009), 1281-1297.

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

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

[20] L.G. Epstein and M. Schneider, Ambiguity, information quality and asset pricing, *J. Finance* 63 (2008), 197-228.

[21] L.G. Epstein and K. Seo, Symmetry of evidence without evidence of symmetry, *Theor. Econ.* 5 (2010), 313-368.

[22] L.G. Epstein and K. Seo, Symmetry or dynamic consistency?, *B.E. J. Theor. Econ.* (Advances), 2011.

[23] L.G. Epstein and K. Seo, A Central Limit Theorem for Capacities, in preparation, 2011b.

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

[25] A. Galichon and M. Henry, Set identification in models with multiple equilibria, *Rev. Econ. Stud.* 78 (2011), 1261-98.

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

[27] I. Gilboa, A. Postelwaite and D. Schmeidler, Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality, *Synthese* 187 (2012), 11-31.

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

[29] F. Gul and W. Pesendorfer, Expected uncertain utility theory, 2010.

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

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

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

[33] B. Jovanovic, Observable Implications of models with multiple equilibria, *Econometrica* 57 (1989), 1431-1437.

[34] M. Kasy, Partial identification, distributional preferences, and the welfare ranking of policies, Harvard, 2012.

[35] T. Kitagawa, Inference and decision for set identified parameters using posterior lower and upper probabilities, UCL, 2011.

[36] P. Klibanoff, S. Mukerji and K. Seo, Relevance and symmetry, 2011.

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

[38] D. Kreps and E.L. Porteus, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46 (1978), 185-200.

[39] Y. Liao and W. Jiang, Bayesian analysis in moment inequality models, *Ann. Stat.* 38 (2010), 275-316.

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

[41] C.F. Manski, *Partial Identification of Probability Distributions*, Springer, 2003.

[42] C.F. Manski, Choosing treatment policies under ambiguity, *Annual Rev. Econ.* 3 (2011), 25-49.

[43] C.F. Manski, Status quo deference and policy choice under ambiguity, *J. Institutional and Theor. Econom.* forthcoming, 2012.

[44] 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.

[45] K. Menzel, Robust decisions for incomplete structural models of social interactions, NYU, 2011.

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

[47] H.R. Moon and F. Schorfheide, Bayesian and frequentist inference in partially identified models, *Econometrica* 80 (2012), 755-82.

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

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

[50] J. Savage, *The Foundations of Statistics*, Dover, New York, 1972.

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

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

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

[54] J. Stoye, New perspectives on statistical decisions under ambiguity, *Annual Rev. Econ.* 4 (2012), 257-82.

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

[56] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010), 167–195.

[57] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003), 147-165.