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

Larry G. Epstein       Kyoungwon Seo

October 16, 2011

Abstract

We consider an individual whose theory of her environment is incomplete and who therefore is concerned that data are correlated and heterogeneous in some unknown fashion. We provide a unified normative axiomatic model of both inference and choice for such an individual. A prime example is an analyst or 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. The central component of the model is a generalization of de Finetti’s exchangeable Bayesian model to accommodate ambiguity. The broad message of the paper is that ambiguity aversion can be fruitfully applied to partially identified models.

*Boston University, lepstein@bu.edu and Department of Managerial Economics and Decision Sciences, Northwestern University, k-seo@kellogg.northwestern.edu. We gratefully acknowledge the financial support of the National Science Foundation (awards SES-0917740 and SES-0918248) and discussions with Marc Henry and Hiro Kaido. We also thank seminar participants at Harvard, Warwick and Northwestern. 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. This version is PRELIMINARY and INCOMPLETE.
1. Introduction

1.1. Motivation and objectives

The opening sentence in Ibragimov and Muller (2010) perfectly summarizes the motivation underlying this paper. They write: “Empirical analyses in economics often face the difficulty that the data are correlated and heterogeneous in some unknown fashion.” They are referring to statistical decision-makers who must decide how to draw inferences from realized data. But a similar difficulty faces an economic decision-maker who must choose an action when payoffs depend on the future outcomes of a number of experiments or random events. A sophisticated decision-maker might be concerned that she does not know precisely the probability law governing these payoff-relevant outcomes and, in particular, that they may be correlated and heterogeneous in some unknown fashion. In both settings, the qualifier “unknown” is crucial. Where the nature of correlation and heterogeneity is known, both can in principle be readily taken into account in both inference and choice. This paper provides normative axiomatic foundations for both inference and choice where unknown correlation and heterogeneity are a concern.

Before elaborating on the model, we describe a concrete example where the noted concern seems natural. The example is drawn from the literature in econometrics addressing models where relevant parameters are only partially identified (see Tamer (2010) for a survey; more references will be given below). Partially identified models arise, for instance, in entry games with multiple equilibria such as have been studied in empirical industrial organization. The following game, due to Jovanovic (1989), illustrates the main points most simply and serves as our running example through the paper.

There are many markets, indexed by \(i = 1, 2, \ldots\). In the \(i^{th}\) market, two firms, \(j = 1, 2\), play the game described by the payoff matrix shown.

<table>
<thead>
<tr>
<th></th>
<th>\text{out}</th>
<th>\text{in}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{out}</td>
<td>0, 0</td>
<td>(0, -\epsilon_{i2})</td>
</tr>
<tr>
<td>\text{in}</td>
<td>(-\epsilon_{i1}, 0)</td>
<td>(\eta^{1/2} - \epsilon_{i1}, \eta^{1/2} - \epsilon_{i2})</td>
</tr>
</tbody>
</table>

The parameter \(\eta\) lies in \((0, 1]\) and the \(\epsilon_{ij}\)’s are observed by players but not by the analyst. She views \(\eta\) as common across markets and the \(\epsilon_{ij}\)’s as uniformly distributed on \([0, 1]\) for each \(i\) and i.i.d. across markets. Her objective is to learn what she can from observed outcomes in a sample of markets and then, in her second role as policy-maker, to make a policy (e.g., regulatory) choice for remaining markets. The costs and benefits of policy options depend on the number
of entrants in those markets and thus the prediction of future entry decisions is important. The analyst’s theory is that a pure strategy Nash equilibrium is played in each market. Every pure strategy Nash equilibrium has either two or zero entrants. Denote by $B$ the outcome where both firms enter and by $N$ the outcome where neither firm enters. Then the set of Nash equilibria is given by:

$$\begin{array}{|c|c|}
\hline
\{B, N\} & \text{if } 0 \leq \epsilon_{11}, \epsilon_{12} \leq \eta^{1/2} \\
\{N\} & \text{otherwise} \\
\hline
\end{array}$$

(1.1)

The multiplicity of equilibria poses difficulties for the analyst because she has no theory of selection. She may believe that such a theory 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, the analyst cannot assign a probability to $N$ being selected in any given market, and she does not understand how selection may differ or be correlated across markets.

This lack of understanding of selection has two important consequences. First it affects inference about $\eta$ because a given sample can be interpreted in more than one way. Outcomes where no one enters the market may indicate that $\eta$ is small, or else that $\eta$ is large (consistent with multiple equilibria) and that $N$ is selected. Consequently, even an infinite sample may not permit identifying a unique value for $\eta$. For example, an infinite sample with empirical frequency of $B$ equal to $q$ is consistent with any value of $\eta$ in the interval $[q, 1]$, the identified set for this sample. The second consequence is that prediction and choice are made more difficult. This is because even given knowledge of $\eta$, the probability of the outcome $B$ in remaining markets could lie anywhere in the interval $[0, \eta]$, and there is no basis for being confident in any single number in that interval for the purpose of prediction and ultimately of choice. In other words, ignorance of the selection mechanism implies a set of likelihoods. When, as here, a theory does not support probabilistic predictions even given knowledge of all its parameters, we refer to it as an incomplete theory.

There are two features of this example that we emphasize because they ‘explain’ our modeling choice. Uncertainty about the $\epsilon_{ij}$’s is probabilistic while uncertainty about selection is deeper and qualitatively distinct. This recalls the distinction between risk and “Knightian uncertainty” or ambiguity. Accordingly, we model the analyst as an ambiguity averse decision-maker. The second feature is that two different factors underlie outcomes across markets - factors that are common to all markets (here $\eta$) and poorly understood idiosyncratic factors.
(here selection) that are the reason for concern about correlation and heterogeneity. The distinction between common and idiosyncratic factors plays an important role in the intuition for our central axiom and in the interpretation of the resulting functional form for utility.

The paper proceeds as follows. In the remainder of this introduction we outline the model and discuss some related literature. Section 2 introduces belief functions and also belief function utility which replaces SEU as the framework within which we explore ‘exchangeability.’ Section 3 presents the axiomatic foundations of our model of choice. The implied representation of utility, corresponding to (1.3), is described in Section 4. The model is extended in Section 5 to incorporate inference. Throughout we illustrate and motivate the formal analysis by relating it to the Jovanovic entry game and to partial identification more generally. In Section 6, we consider the problem of prediction of frequencies within the entry game and highlight the effects of ambiguous correlation and heterogeneity. The concluding section outlines an extension to more general entry games and considers additional related literature. Proofs are collected in appendices.

1.2. Model outline

Consider a sequence of experiments and choice between bets (or acts) on their outcomes. In the entry game example, each market corresponds to an experiment and alternative policies can be thought of as different bets on market outcomes (the number of entrants). The dominant model of choice in such a setting is the exchangeable Bayesian model, according to which the policy maker maximizes subjective expected utility (SEU) under the assumption that selection is i.i.d. across markets. The selection probability for $B$ is unknown but the analyst has a probability distribution to represent her uncertainty. In the more general case of repeated experiments, she uses a predictive prior $P$ having the “conditionally i.i.d.” form

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

where: $S$ is the set of outcomes of any single experiment; $\ell \in \Delta(S)$ is a generic probability law for any single experiment; $\ell^\infty$ denotes the corresponding i.i.d. product measure on $S^\infty$; and $\mu$ represents beliefs (either a prior or posterior after updating) on $\Delta(S)$. Foundations for this model are well known: Savage (1972) and Anscombe and Aumann (1963) axiomatize the SEU specification and de Finetti (1937) shows that the conditionally i.i.d. form for the predictive prior is
characterized by 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. Expressed in terms of behavior, this is the assumption that there is indifference between any two bets that differ only by a permutation or reordering of experiments. Such indifference is natural (even compelling) in the Jovanovic entry game where there is no reason to distinguish between markets.\footnote{See Section 7 for more general entry games where markets are distinguished by characteristics observable by the analyst. Then symmetry is again natural after correcting for these differences.} Besides its elegance and the normative appeal of its axiomatic underpinnings, another strength of the model is that Bayesian updating (applied to $\mu$) renders the model dynamic and delivers dynamic consistency. Thus extended, the exchangeable Bayesian model is the canonical model of learning or inference in economics and Bayesian statistics.

However, in spite of its strengths, the exchangeable Bayesian model fails to capture the intuition described for the entry game. This is suggested by the representation (1.2): though the beliefs in (1.2) reflect uncertainty about the true selection probability, they also indicate \textit{certainty} that the selection mechanism is i.i.d. across markets thus excluding any concern about correlation and heterogeneity.\footnote{This critique will be expressed in behavioral terms below.} The difficulty with the model is not due to the exchangeability assumption. Rather we argue that the SEU framework, specifically the Independence Axiom, is what prevents the model from capturing intuitive features of the entry game.

To permit a role for ambiguity, we consider preference over acts that conforms to Choquet expected utility (Schmeidler (1989)) where the capacity is a belief function - we call this model \textit{belief function utility}.\footnote{See Section 2 and Appendix A for more on belief functions and the corresponding utility functions.} Belief function utility is also a special case of the multiple-priors model (Gilboa and Schmeidler (1989)). Using it as the basic framework, we impose two simple axioms - Symmetry (corresponding to de Finetti’s exchangeability) and Weak Orthogonal Independence (relaxing the Independence Axiom). These axioms are shown (Theorem 4.1) to characterize the following de Finetti-style representation for the belief function $\nu$ on $S^\infty$:

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

where $Bel(S)$ denotes the set of all belief functions on $S$, $\mu$ is a probability measure on $Bel(S)$, and $\theta^\infty$ denotes a suitable “i.i.d. product” of the belief function $\theta$. 

\footnote{1See Section 7 for more general entry games where markets are distinguished by characteristics observable by the analyst. Then symmetry is again natural after correcting for these differences.}
The de Finetti-Savage model is the special case where each $\theta$ in the support of $\mu$ is additive.

Besides this formal similarity to the Bayesian exchangeable model, there are three other important similarities. First, by adapting Epstein and Seo (2010), we provide (Section 5) an axiomatically founded model of updating to go along with the static representation (1.3). Updating consists of a suitable application of Bayes’ rule to the prior $\mu$. The unified foundations for both inference and choice in the context of partially identified models is a major contribution. Second, the axiomatic foundations for the comprehensive (that is, including updating) model are arguably simple and normatively appealing. Finally, the model is tractable. Thus, for example, results in Moon and Schorfheide’s (forthcoming) Bayesian analysis of inference in partially identified models apply, as do results in Acemoglu et al. (2009).\footnote{See Section 5.} Tractability of the choice component is demonstrated through an application to the problem of predicting the empirical frequency of the dual entry outcome in the Jovanovic entry game (see Section 6). An important difference from the exchangeable Bayesian model is that our model captures the intuitive behavior for the entry game given concern with unknown correlation and heterogeneity due to the multiplicity of equilibria.

Finally, we clarify the way in which incompleteness of the theory and thus partial identification are reflected in our model. They are reflected in choice through the non-Bayesian model of preference. While making (predictions or) choices, the decision-maker protects herself from the incompleteness of her theory by considering the worst scenario. However, inference is Bayesian. Therefore, some elaboration of this component is needed. The key to inference is the nature of the likelihood function. In the entry game, our model yields (after suitable translation) a likelihood function of the form

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

where each $\lambda_\eta$ is a probability measure over selection mechanisms (represented by $q$) that might be thought of as describing subjective uncertainty about which i.i.d. selection mechanism is valid. The fact that there is uncertainty about the selection probability even given $\eta$ reflects the incompleteness of Nash equilibrium as a theory in the absence of an accompanying theory of selection. Similar likelihood specifications are used without foundations in Acemoglu et al. (2009) and in
Moon and Schorfheide (forthcoming). We provide axiomatic foundations for such
likelihoods, not just in an entry game context, but more generally in an abstract
setting where the decision-maker’s theory is suitably incomplete. In the general
setting, the parameter \( \eta \) is replaced by a belief function over a single experiment
(the set \( S \) of outcomes), and it is the belief function that serves as the partially
identified parameter. Section 5 elaborates.

1.3. Related literature

Two strands of literature are particularly relevant to the applied aspect of this
paper.\(^5\) The first is robust Bayesian statistics (see Berger (1994) for a survey
and the collection of papers in Rios and Ruggeri (2000)). In general, robustness is
sought to possible misspecifications of the prior, likelihood and the loss (or utility)
function. However, the overwhelming majority of papers address robustness with
respect to the prior. Our model of inference can be seen as yielding robustness
with respect to the typical i.i.d. assumption for likelihoods or the data generating
mechanism, the need for which is illustrated by the entry game. Robustness of
this sort is the concern of the frequentist analysis in Ibragimov and Muller (2010).
However, we have not found any Bayesian statistical models addressing it.

Another important strand of literature concerns partial identification. Tamer
(2003, 2010), Beresteau et al. (forthcoming) and Galichon and Henry (2011),
in particular, emphasize the importance of focussing on testable implications of
the core economic theory alone, even where that prevents point identification of
key parameters because the theory is incomplete. Entry games with multiple
equilibria are a prime example and area of application. The latter two papers
exploit random set theory and belief functions to characterize the identified set of
parameters, which connects them to this paper in terms of tools used; however,
the questions asked differ. For applications see, for example, Ciliberto and Tamer
(2009), Bajari et al. (2010) and Jia (2005), and for estimation and inference in
partially identified models see Chernozukhov et al. (2007) and Beresteau and
Molinari (2008); they all adopt a frequentist approach.

More closely related to the present paper are the Bayesian approaches to in-
fERENCE in Moon and Schorfheide (2011) and Liao and Jang (2010), and the ro-
bust Bayesian analysis in Kitagawa (2011), where robustness is with respect to a
misspecified prior. To varying degrees, these papers either adopt ad hoc (albeit
possibly intuitive and tractable) specifications and do not address robustness with

\(^5\) Connections to the decision theory literature will be discussed in the concluding section.
respect to correlation and heterogeneity. Also related is the literature on choice in partially identified models. For example, see Manski (2011) and the references therein, and Kasy (2011). 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. Kitagawa (2011) models statistical choice. More broadly, decisions are the ultimate objective of most statistical analyses. Models that study inference without specifying an accompanying model of choice beg the question whether an appropriate model of choice can be added. This is not always clear because, as noted above, partial identification creates difficulties for modeling decision-making. Our model is axiomatically well-founded and unifies inference and choice.

2. Belief Function Utility

We define belief functions on an abstract state space $\Omega$. The presentation is intended to be broadly accessible rather than ‘efficient.’ Appendix A contains many supporting references and technical details and verifies the consistency of our definitions with other common formulations.

Think of states in $\Omega$ as representing the payoff relevant uncertainty facing a decision-maker. That is, the outcome of any chosen physical action depends on the realized state in $\Omega$. Accordingly, choice is modeled formally as the choice between acts $f$ (see (2.1)). Thus beliefs $\nu$ on $\Omega$ are important. Formulating beliefs on $\Omega$ is a challenge because the decision-maker’s understanding, or theory, of $\Omega$ is incomplete. Her theory is based on an auxiliary state space $\hat{\Omega}$ where her understanding permits her to assign probabilities using a measure $m$. The auxiliary space is simpler than $\Omega$, or provides a coarse picture of $\Omega$, in that each point in $\hat{\Omega}$ corresponds to a set of points in $\Omega$, that is, there is a correspondence $\Gamma$ from $\hat{\Omega}$ into $\Omega$.

\[
(\hat{\Omega}, m) \xrightarrow{\Gamma} (\Omega, \nu) \downarrow_f [0, 1]
\]

(2.1)

Awareness of this coarseness and a conservative attitude lead to beliefs on $\Omega$
represented by \( \nu \) given by\(^{6}\)
\[
\nu (A) = m \left( \{ \widehat{\omega} \in \widehat{\Omega} : \Gamma (\widehat{\omega}) \subset A \} \right).
\] (2.2)

Any function \( \nu \) on the Borel \( \sigma \)-algebra of \( \Omega \) that can be constructed in this way is called a belief function.\(^{7}\) Refer to \((\widehat{\Omega}, m, \Gamma)\) as representing, or generating, \( \nu \).

Next define the corresponding utility function. The objects of choice are (Borel measurable) acts \( f : \Omega \rightarrow [0, 1] \), which for simplicity are restricted to have finite range (such acts are commonly called ‘simple’). The utility of any such act \( f \) is computed by
\[
U (f) = \int_{\widehat{\Omega}} \left( \inf_{\omega \in \Gamma (\widehat{\omega})} f (\omega) \right) dm (\widehat{\omega}).
\] (2.3)

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.

Two observations help to further clarify the functional form (2.3) for utility. First, it is a special case of the multiple-priors model (Gilboa and Schmeidler (1989)) because it is readily shown that (2.3) can be expressed alternatively as
\[
U (f) = \min_{P \in \text{core} (\nu)} \int_{\Omega} f dP,
\] (2.4)
where the core of \( \nu \) is defined by
\[
\text{core} (\nu) = \{ P \in \Delta (\Omega) : P (\cdot) \geq \nu (\cdot) \},
\]
and is given by\(^{8}\)
\[
\text{core} (\nu) = \left\{ P \in \Delta (\Omega) : P = \int_{\widehat{\Omega}} p_{\widehat{\omega}} dm (\widehat{\omega}), p_{\widehat{\omega}} \in \Delta (\Gamma (\widehat{\omega})) \text{ m-a.e.} \right\}.
\] (2.5)

\(^{6}\)More precisely, \( \Omega \) and \( \widehat{\Omega} \) are compact metric spaces, \( m \) is a Borel probability measure, the correspondence \( \Gamma \) is measurable and nonempty-compact-valued, and \( \nu \) is defined on the Borel \( \sigma \)-algebra of \( \Omega \). Finally, outcomes lie in the unit interval because they are denominated in utils (see further explanation below).

\(^{7}\)If \( \Gamma \) is singleton-valued and hence a random variable, then \( \nu \) is a probability measure and (2.2) is the familiar formula for computing induced distributions.

\(^{8}\)When the support of \( m \) is not finite, a measurability assumption for \( \widehat{\omega} \mapsto p_{\widehat{\omega}} \) must be added to give meaning to this expression.
It is also a special case of Choquet expected utility (Schmeidler (1989)) because utility can be expressed alternatively as a Choquet integral in the form

\[ U(f) = U_\nu(f) = \int_\Omega f \, d\nu, \]  

(2.6)

where \( \nu \) is defined by (2.2). Any one of the equivalent equations (2.3), (2.4) or (2.6) defines belief function utility.

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

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

Let \( \theta \) be a belief function on \( S \) describing beliefs about any single experiment. We define its i.i.d. product, denoted \( \theta^\infty \), representing beliefs about the sequence of experiments, including how they are related and differ. Let \( \theta \) be generated by the triple \((\hat{S}, m, \Gamma)\) and consider the triple \((\hat{\Omega}, m^\infty, \Gamma^\infty)\), where: \( \hat{\Omega} = \left( \hat{S} \right)^\infty \), \( m^\infty \) is the ordinary i.i.d. product of the probability measure \( m \), and \( \Gamma^\infty \) is the correspondence \( \Gamma^\infty : \hat{\Omega} = \left( \hat{S} \right)^\infty \leadsto \Omega = S^\infty \) given by

\[ \Gamma^\infty (\hat{s}_1, \hat{s}_2, \ldots) = \Gamma (\hat{s}_1) \times \Gamma (\hat{s}_2) \times \ldots \]  

(2.7)

Then \( \theta^\infty \) is the belief function on \( \Omega \) generated as in (2.2) by \((\hat{\Omega}, m^\infty, \Gamma^\infty)\). It models a form of stochastic independence between experiments. In particular, the following ‘product property’ is readily verified:

\[ \theta^\infty (A_I \times A_J \times S^\infty) = \theta^\infty (A_I \times S^\infty) \theta^\infty (A_J \times S^\infty), \]  

(2.8)

for \( A_I \subset \Pi_{i \in I} S_i \) and \( A_J \subset \Pi_{j \in J} S_j \) where \( I, J \subset \mathbb{N} \) are finite and disjoint.

As an illustration, fix an event \( E \subset S \) and define \( \theta \) on \( S \) by

\[ \theta(A) = \begin{cases} 1 & A \supset E \\ 0 & \text{otherwise} \end{cases} \]

9Appendix A shows that \( \theta^\infty \) is well defined (if \((\hat{S}, m, \Gamma)\) and \((\hat{S}', m', \Gamma')\) both generate \( \theta \), then they both lead to the same belief function on \( S^\infty \)) 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).
Then $\theta$ is a belief function and it captures certainty that $E$ will occur, but ignorance within $E$; the corresponding core, $\text{core}(\theta)$, equals the entire probability simplex $\Delta(E)$. The core of $\theta^\infty$ is the entire simplex $\Delta(E^\infty)$. It obviously contains many measures that are not product measures (giving scope to model ambiguity about correlation) and also many product measures that are not identical products (thus permitting ambiguity about heterogeneity). The interpretation of (2.8) as a form of stochastic independence in conjunction with the reference to (ambiguous) correlation may seem contradictory. However, as demonstrated by Hendon et al. (1996) and Ghirardato (1997), stochastic independence is multi-faceted when there is ambiguity - there are degrees or types of independence, and the kind reflected by (2.8) is weak enough to permit capturing also ambiguity about correlation. Moreover, this feature is not restricted to this example. It is a general feature that, for any (non-additive) belief functions $\theta$ on $S$, 

$$
\text{core}(\theta^\infty) \supsetneq \text{core}(\theta) \otimes \text{core}(\theta) \otimes \ldots \\
\equiv \{p_1 \otimes p_2 \otimes \ldots : p_i \in \text{core}(\theta) \text{ for every } i\}.
$$

We conclude this preliminary section by describing how belief functions and their i.i.d. products arise naturally in the entry game.

Entry game example: The uncertainty facing the analyst or policy-maker is the number of entrants in each market. Thus uncertainty about any single market is represented by the state space $S = \{B, N\}$. Each given parameter $\eta$ induces a belief function on $S$. This is through the equilibrium correspondence $\Gamma_\eta$,

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

defined as in (1.1), and the assumed uniform distribution $m$ on $[0, 1]^2$. The triple $([0, 1]^2, m, \Gamma_\eta)$ induces the belief function $\theta_\eta$, where

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

The set of priors corresponding to $\theta_\eta$ is the set of all measures on $\{B, N\}$ for which the probability of $B$ lies in the interval $[0, \eta]$. In other words, $\text{core}(\theta_\eta)$ can be

\footnote{Recall that $B$ denotes the outcome where both firms enter and $N$ the outcome where neither enters.}
identified with the latter interval. The i.i.d. product of $\theta_\eta$ is constructed in the similar way given that the $\epsilon_i$’s are assumed to be distributed i.i.d. across markets, using the equilibrium sequence correspondence $\Gamma^\infty_\eta$, where

$$\Gamma^\infty_\eta (\epsilon_1, \epsilon_2, \ldots) = \Gamma_\eta (\epsilon_1) \times \Gamma_\eta (\epsilon_2) \times \ldots$$

Some insight into the core of $(\theta_\eta)^\infty$ is gleaned from (2.5). For $P$ a measure on $S^\infty$, denote by $mrg_{\{1,2\}} P$ the marginal on $S_1 \times S_2$. Then

$$\{mrg_{\{1,2\}} P : P \in \text{core } ((\theta_\eta)^\infty)\} =$$

$$(1 - \eta)^2 \Delta (\{N_1 \times N_2\}) + \eta^2 \Delta (\{B_1, N_1\} \times \{B_2, N_2\})$$

$$\eta (1 - \eta) [\Delta (\{B_1, N_1\} \times \{N_2\}) + \Delta (\{N_1\} \times \{B_2, N_2\})].$$

This set of probability mixtures attaches weight $\eta^2$ to the set of all probability measures on $\{B_1, N_1\} \times \{B_2, N_2\}$, including both nonproduct measures and nonidentical products, thus indicating that both correlation and heterogeneity ambiguity are reflected.

The above specification is what we have in mind below when we apply our model to the ‘entry game example.’ However, it is important to emphasize that the preceding describes only one way to construct a belief function for each $\eta$. The construction above assumes complete ignorance about selection and this leads to any number in the interval $[0, \eta]$ being a possible probability for $B$. However, beliefs are subjective and unobservable as is the analyst’s underlying theory of selection. The definition of belief functions permits a host of alternative theories of selection through alternative measures $m$, ($m$ is after all subjective), and more fundamentally through alternative auxiliary spaces and correspondences. Thus, for example, our model permits the probability interval corresponding to the parameter $\eta$ to be a strict subset of $[0, \eta]$. Indeed, any interval defines a unique belief function. For example, if the interval is $[p, \overline{p}]$, take the auxiliary state space $\tilde{S} = \{\{B\}, \{N\}, \{B, N\}\}$, the probability measure $m$ given by $m (\{B\}) = p$, $m (\{N\}) = 1 - \overline{p}$ and $m (\{B, N\}) = \overline{p} - p$, and the correspondence $\Gamma$ given by $\Gamma (\{B\}) = \{B\}$, $\Gamma (\{N\}) = \{N\}$ and $\Gamma (\{B, N\}) = \{B, N\}$. Then (2.2) defines the belief function $\theta$ for which

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

This generality can be seen as a strength, or alternatively as a weakness because of the degrees of freedom left for the decision-maker in trying to formulate beliefs. (See, however, Corollary 4.2 and the preceding discussion.)
A separate lesson that can be drawn from the preceding is that for binary state spaces a belief function can be thought of simply as a probability interval.

The outline thus far has proceeded as though the decision-maker knew $\eta$. In fact, there will typically be uncertainty about $\eta$, as indicated in (1.3). Before considering beliefs and utility more closely, we first describe the axiomatic foundations for our model.

3. Foundations

As described in the previous section, we consider a state space of the form

$$\Omega = S_1 \times S_2 \times \ldots = S^\infty,$$

where $S$, the set of possible outcomes for each experiment, is compact metric. Objects of choice are (Borel measurable and simple) acts $f: \Omega \to [0, 1]$. 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, without loss of generality, 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.

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.

The set of all acts is $\mathcal{F}$. We study preference $\succeq$ on $\mathcal{F}$. Three axioms are imposed.

**Axiom 1 (Belief Function Utility).** The preference $\succeq$ admits representation by a belief function utility $U$.

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, expressed by the next two axioms, that is intuitive given repeated experiments. The 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 $\Omega$ be the set of (finite) permutations on $\mathbb{N}$. For $\pi \in \Pi$ and $\omega = (s_1, s_2, \ldots) \in \Omega$, let $\pi \omega = (s_{\pi(1)}, s_{\pi(2)}, \ldots)$. Given an act $f$, define the permuted act $\pi f$ by $(\pi f)(s_1, \ldots, s_n, \ldots) = f(s_{\pi(1)}, \ldots, s_{\pi(n)}, \ldots)$. For example, if $f = B_1 N_2$ and $\pi$ 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 $\pi$,

$$f \sim \pi f.$$ 

Symmetry is intuitive in situations where, as in the Jovanovic entry game, information about the experiments is symmetric. However, 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. The final axiom leaves room for such situations and accordingly for ambiguity about how experiments are related. 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 final axiom is:\footnote{Talagrand (1978) contains the study of symmetric belief functions, where WOI for the corresponding utility function is not assumed.}

**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 $f \perp g$,

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

The Independence axiom requires the similar invariance of rankings for all (not necessarily orthogonal) acts.\footnote{Because we denominate outcomes in utils, Independence would imply expected value maximization.} 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 \( \eta \), 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

\[
\frac{1}{2}B_1 + \frac{1}{2}(.8N_1) \succ B_1,
\]

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.

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

\[
\frac{1}{2}B_1 + \frac{1}{2}N_1 \succ \frac{1}{2}B_1 + \frac{1}{2}N_2.
\]

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 (3.2).

Finally, 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, their mixture is strictly preferable if unknown correlation is a concern, that is, one expects the rankings

\[
\frac{1}{2}\{B_1B_2, N_1N_2\} + \frac{1}{2}\{B_1B_3, N_1N_3\} \succ \{B_1B_2, N_1N_2\} \sim \{B_1B_3, N_1N_3\}.
\]
favor the bet \( \{B_1B_2, N_1N_2\} \), or the variable may be similar in markets one 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 (3.1)-(3.3), but it remains to interpret WOI and to determine what kind of behavior it does exclude. 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 possible connection is that market outcomes depend on the common factor \( \eta \). If there is ambiguity about \( \eta \), 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 \( \eta \). It also 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. In that case, one would expect that (with rescaling as in (3.1)),

\[
\frac{1}{2}B_1 + \frac{1}{2}(.8N_2) \succ B_1 \sim .8N_2,
\]

contrary to WOI.

One would like to say more about the kind of 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, as noted in the introduction, “stochastic independence” is multifaceted if there is ambiguity and not well understood behaviorally. However, one can view the axiom as providing such behavioral meaning (given the absence of ambiguity about \( \eta \)) to one form of stochastic independence. Its simple statement is another advantage because it 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

\[
\frac{1}{2}B_1 + \frac{1}{2}N_3 \sim \frac{1}{2}B_2 + \frac{1}{2}N_3,
\]

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.

Finally, it is easy to show that WOI is satisfied if and only if \( U \) satisfies, for all \( \alpha \) and finitely-based and orthogonal acts \( f \) and \( g \),

\[
U(\alpha f + (1 - \alpha) g) = \alpha U(f) + (1 - \alpha) U(g).
\]  (3.4)
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 (2.6), any act $f$ is indifferent to the constant act giving $U(f)$ in every state. Therefore, the expression (3.4) is a meaningful statement about preference.

4. The Representation

We show in this section that the three axioms above characterize a “conditionally i.i.d.” representation analogous to de Finetti’s. The belief-function utility $V$ on $\mathcal{F}$ is called an i.i.d. (belief-function) utility if there exists $\theta$, a belief function on $S$, such that

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

When $\theta$ is additive, the function reduces to expected value with an i.i.d. probability measure. Just as de Finetti shows that mixtures of i.i.d. probability measures generate all exchangeable measures, we show that mixtures of i.i.d. belief function utilities generate all utility functions satisfying our axioms.

Denote by $\text{Bel}(S)$ the set of belief functions on $S$;\textsuperscript{13} a generic element is denoted $\theta$.

**Theorem 4.1.** Let $\succeq$ be a preference order on the set of acts $\mathcal{F}$. Then the following statements are equivalent:

(i) $\succeq$ satisfies Belief Function Utility, Symmetry and Weak Orthogonal Independence.

(ii) There exists a (necessarily unique) Borel probability measure $\mu$ on $\text{Bel}(S)$ such that $\succeq$ is represented by $U$, where

$$U(f) = \int_{\text{Bel}(S)} V_{\theta^\infty}(f) d\mu(\theta), \text{ for every } f \in \mathcal{F}. \quad (4.1)$$

(iii) There exists a (necessarily unique) Borel probability measure $\mu$ on $\text{Bel}(S)$ such that $\nu$, the belief-function provided by Belief Function Utility, can be expressed in the form

$$\nu(A) = \int_{\text{Bel}(S)} \theta^\infty(A) d\mu(\theta), \text{ for every Borel } A \subset \Omega. \quad (4.2)$$

\textsuperscript{13}Endow $\text{Bel}(S)$ with the topology for which $\theta_n \to \theta$ if and only if $\int fd\theta_n \to \int fd\theta$ for every continuous function $f$ on $S$, where the integral is in the sense of Choquet. Then $\text{Bel}(S)$ is compact metric.
Entry game example (continued): Interpret the functional form in the context of the entry game example, where \( S = \{B, N\} \). Like de Finetti’s representation, (4.2) also admits a “conditionally i.i.d.” interpretation. However, an important difference is that each market is described by a belief function \( \theta \) rather than by a single probability. Because each belief function \( \theta \) corresponds to an interval \([\theta(B), 1 - \theta(N)]\) of probabilities for \( B \) (dual entry), even given certainty about the correct belief function \( \theta \) there would remain ambiguity about the selection probability in each market. As indicated, the decision-maker perceives each market as being described by the same interval. A concern with possible heterogeneity can be understood as arising because she fears that different probabilities from that interval could apply to different markets. Correlation across markets could arise similarly. At a behavioral level, the functional form accommodates the rankings (3.1)-(3.3) that illustrate our intuition for the entry game. The first two rankings are immediate. For the third, let \( \mu \) attach positive probability to \( \theta \), where \( \theta(N) > 0 \) and \( \theta(B) + \theta(N) < 1 \). Abbreviate the bet \( \{B_1B_2, N_1N_2\} \) by \( f \) and let \( \pi \) be the permutation that switches the second the third markets. Then, by straightforward calculations, \( V_{\theta^*}(f) = (\theta(B))^2 + (\theta(N))^2 < V_{\theta^*}(\frac{1}{2}f + \frac{1}{2}\pi f) = (\theta(B))^2 + (\theta(N))^2 + \theta(B)\theta(N)(1 - \theta(B) - \theta(N)) \). Therefore, (3.3) follows.

Finally, focus on the prior \( \mu \). The theorem establishes its existence (given the axioms) but does not provide guidance to the decision-maker as to how to arrive at a prior. However, as we show now the model does provide a way for the decision-maker to calibrate her beliefs. We illustrate this first in the entry game and provide a more general result below.

The primitive uncertainty is about the value of \( \eta \). Suppose the analyst forms a prior over \( \eta \)’s. In the application of the model to the entry game described in Section 2, each \( \eta \) is associated with the probability interval \( I_\eta = [0, \eta] \) and hence with the corresponding belief function \( \theta_\eta \). Thus the prior (also denoted \( \mu \)) over \( \eta \)’s corresponds to a unique prior over belief functions, and we need only consider calibration of beliefs about \( \eta \). The key is the connection to beliefs about empirical frequencies, which arguably are easier for the decision-maker to determine.

For a given (infinite) sample \( \omega \), denote by \( \Psi_\eta(\omega) \) the proportion of the first \( n \) markets where both firms enter. Then a law of large numbers (LLN) for i.i.d. belief functions due to Maccheroni and Marinacci (2005) implies that

\[
(\theta_\eta)^\infty \left( \\{ \omega \in \Omega : [\liminf \Psi_\eta(\omega), \limsup \Psi_\eta(\omega)] \subset I_\eta \} \right) = 1. \tag{4.3}
\]
Further, these bounds on empirical frequencies are tight in the sense that
\[
[a > \theta_\eta(B) = 0 \text{ or } b < 1 - \theta_\eta(N) = \eta] \implies 0 = (\theta_\eta)^\infty \left( \{[\lim\inf \Psi_n(\omega), \lim\sup \Psi_n(\omega)] \subset [a, b]\} \right).
\]

Therefore, the representation (4.2) implies that, for every \(0 \leq b \leq 1\),
\[
\begin{align*}
\mu(\{\eta : I_\eta \subset [0, b]\}) &= U(\{\omega : [\lim\inf \Psi_n(\omega), \lim\sup \Psi_n(\omega)] \subset [0, b]\}) \\
&= U(\{\omega : \lim\sup \Psi_n(\omega) \leq b\}).
\end{align*}
\]

(The right side denotes the utility of the bet on the indicated event.) Finally, take \([0, b] = I_\eta\). Because \(I_\eta \subset I_\eta\) if and only if \(\eta \leq \bar{\eta}\), conclude that
\[
\mu(\{\eta : 0 \leq \eta \leq \bar{\eta}\}) = U(\{\omega : \lim\sup \Psi_n(\omega) \leq \bar{\eta}\}).
\]

In other words, the probability assigned to values of \(\eta\) no greater than \(\bar{\eta}\) 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 \(\bar{\eta} + \epsilon\) in all sufficiently large samples. The decision-maker need “only” decide on such certainty equivalents in order to form a prior \(\mu\).

Only the \(\lim\sup\) appears because we have specified the intervals \(I_\eta\) to have zero as their left endpoint. For other specifications of \(I_\eta\), if \(I_\eta \subset I_\eta \iff \eta \leq \bar{\eta}\), then\(^\text{14}\)
\[
\mu(\{\eta : 0 \leq \eta \leq \bar{\eta}\}) = U(\{\omega : [\lim\inf \Psi_n(\omega), \lim\sup \Psi_n(\omega)] \subset I_\eta\}).
\]

Compare with the situation where the selection probability of \(B\) is known to be \(q\) in each market and i.i.d. across markets. Then the decision-maker maximizes subjective expected utility with an exchangeable predictive prior, the classical LLN for exchangeable measures implies certainty that the empirical frequency of \(B\) converges to \(q\eta\), and beliefs about \(\eta\) and bets about empirical frequencies are related by
\[
\mu(\{\eta : 0 \leq \eta \leq \bar{\eta}\}) = U(\{\omega : \lim \Psi_n(\omega) \leq q\bar{\eta}\}).
\]

The condition (4.7) generalizes this well-known relation to our setting where there does not exist a theory of selection, where there is uncertainty about whether selection is i.i.d. and consequently, where it is not certain that empirical frequencies converge.

\(^{14}\)A more general result follows in the corollary.
The connection between beliefs about parameters and bets on frequencies is a general feature of our model. Return to the general setting of a nonbinary (compact metric) state space \( S \). Denote by \( \Psi_n (\cdot) (\omega) \) the empirical frequency measure given the sample \( \omega \); \( \Psi_n (A) (\omega) \) is the empirical frequency of the event \( A \subset S \) in the first \( n \) experiments. Then the following corollary of our main result can be proven.

**Corollary 4.2.** Let \( U \) be a belief function utility satisfying Symmetry and Weak Orthogonal Independence. Then the equivalent statements in Theorem 4.1 are equivalent also to the following: There exists a probability measure \( \mu \) on \( \text{Bel} (S) \) satisfying both (i) \( \mu \) represents \( U \) in the sense of (4.1); and (ii) for every finite collection \( \{A_1, \ldots, A_J\} \) of Borel subsets of \( S \), and for all \( a_j \leq b_j, j = 1, \ldots, J \),

\[
\mu \left( \bigcap_{j=1}^{J} \{ \theta : [\theta (A_j), 1 - \theta (S \setminus A_j)] \subset [a_j, b_j] \} \right) = U \left( \bigcap_{j=1}^{J} \{ \omega : \lim \inf \Psi_n (A_j) (\omega), \lim \sup \Psi_n (A_j) (\omega) \subset [a_j, b_j] \} \right).
\]

Equation (4.8) relates the prior \( \mu \) over parameters, here belief functions, to the evaluation of bets on empirical frequencies for the events \( A_1, \ldots, A_J \). More precisely, the \( \mu \)-measures of the sets shown are so related. However, Proposition C.3 shows that \( \mu \) is completely determined by its values on these sets.

5. Updating

Thus far we have described the model of choice at any stage. Past observations underlie beliefs but their connection to beliefs is suppressed. Now we focus explicitly on this connection. Thus, for example, we describe how beliefs about (and preferences over bets on) remaining markets are updated in response to observation of the outcomes in some markets. We do so by adapting the model of updating in Epstein and Seo (2010, Section 6). We briefly recall the main elements of that model, but we refer the reader to that paper for elaboration. Henceforth, \( S \) is assumed to be finite.

To describe dynamic choice we adopt as primitives not only an ex ante preference \( \succeq_0 \) but also conditional preference \( \succeq_{s^n} \) at each node, where \( s^n = (s_1, \ldots, s_n) \) is the history at stage \( n \). All preferences are defined on \( \mathcal{F} \), the set of simple
acts. We assume that each preference satisfies the axioms in Theorem 4.1 and thus admits representation as in (4.1). Denote by $\mu_n(\cdot \mid s^n)$ the (unique) measure representing $\succeq_{n,s^n}$, or the posterior after observing $s^n$. The issue is how to arrive at posteriors.

As is well-known, the difficulty in updating ambiguity averse preferences is the tension with dynamic consistency.\textsuperscript{15} We relax this tension by requiring dynamic consistency in a limited, but still interesting class of environments, namely, where an individual first samples and observes the outcomes $s^n = (s_1, \ldots, s_n)$ of $n$ experiments, and then chooses how to bet on the outcomes of remaining experiments. In particular, updating occurs only once, and each experiment serves either as a signal or is payoff relevant, but not both. The decision-maker does not bet on the outcomes of the first $n$ experiments - they serve purely as signals - and she does not update further after observing outcomes of any experiments beyond the $n^{th}$ - they only determine payoffs to previously chosen bets. Accordingly, we assume that, for any $n$ and for any acts $f'$ and $f$ that depend only on the outcomes of experiments $n + 1$ and on,

$$f' \succeq_{n,s^n} f \quad \text{for all } s^n \implies f' \succeq_0 f$$

and

$$f' \succ_{n,s^n} f \quad \text{for some } s^n \implies f' \succ_0 f.$$ 

Refer to this property as \textit{Weak Dynamic Consistency (WDC)}.

We adopt two additional axioms. The first is \textit{Consequentialism}, stating that at any node $(n, s^n)$, unrealized parts of the tree do not matter: For all acts $f'$ and $f$, $f' \sim_{n,s^n} f$ if $f'(s^n, \cdot) = f(s^n, \cdot)$. The second requires that the inference drawn from a sample $s^n$ not depend on the order in which its outcomes were realized: $\succeq_{n,s^n} = \succeq_{n,\pi s^n}$ for all finite permutations $\pi$. Call this property \textit{Commutativity}.

**Theorem 5.1.** Let the ex ante and conditional preferences $\succeq_0$ and $\{\succeq_{n,s^n}\}$ be represented as in Theorem 4.1 by the prior $\mu_0$ and the posteriors $\{\mu_n(\cdot \mid s^n)\}$. Then preferences satisfy Weak Dynamic Consistency, Consequentialism and Commutativity if and only if there exists a likelihood function $L : Bel(S) \rightarrow \Delta(S^\infty)$ such that: (i) each $L(\cdot \mid \theta)$ is exchangeable, that is,

$$L(\cdot \mid \theta) = \int_{\Delta(S)} q^\infty(\cdot) d\lambda (q) \quad (5.1)$$

\textsuperscript{15}For a particularly stark expression of this tension see Epstein and Seo (2011).
for some probability measure $\lambda_0$ on $\Delta(S)$; and (ii) $\mu_n(\cdot \mid s^n)$ is obtained by applying Bayes’ rule to the prior $\mu_0$ and the likelihood $L$.

The updating rule described here constitutes the inference component of our model. The fact that it is Bayesian has the advantage that results from Bayesian learning theory translate directly. As noted in the introduction, the structure of the likelihood function is important in reflecting incompleteness of the individual’s theory. The measure $\lambda_0$ suggests uncertainty even given $\theta$. Because the likelihood $L$ is used for inference (as opposed to prediction), it is as if the individual is uncertain what any given realized sample reveals about $\theta$. Difficulty in interpreting signals is the intuitive motivation given by Acemoglu et al. (2009) for their adoption of such likelihood functions (in their case $\theta$ is an abstract parameter rather than a belief function). They show that such uncertainty about signals can cause posteriors to fail to converge to certainty about a single $\theta$, that is, it can lead to partial identification. In our setting, if $S$ is binary and thus each belief function $\theta$ corresponds to a probability interval $I_\theta$, failure of full identification can occur when intervals overlap ($I_\theta \cap I_{\theta'} \neq \emptyset$ and $\theta \neq \theta'$). We illustrate this further in the entry game example.

A final comment about updating in the general setting is that the theorem does not restrict the likelihood function other than by exchangeability. Any exchangeable likelihood delivers the noted axioms. As illustrated next in the entry game, some additional restrictions seem natural, but they are not implied by the axioms.

**Entry game example (continued):** As pointed out earlier, each $\eta$ corresponds to a unique probability interval $I_\eta = [0, \eta]$ and to a unique belief function $\theta_\eta$ on $S = \{B, N\}$. Primitive uncertainty about $\eta$ induces the prior over belief functions. Therefore, (5.1) takes the form

$$L(\cdot \mid \theta_\eta) = \int_{\Delta(S)} q^\infty(\cdot) \, d\lambda_{\theta_\eta}(q),$$

or, with the obvious change of notation,

$$L(\cdot \mid \eta) = \int_{\Delta(S)} q^\infty(\cdot) \, d\lambda_{\eta}(q).$$

An additional restriction that suggests itself is that $\lambda_\eta$ have support contained in $\text{core}(\theta_\eta) = \{q \in \Delta(S) : q(B) \leq \eta\}$. One might also impose that it has
full support on \( \text{core}(\eta) \). For example, \( \lambda_{\eta} \) could be the uniform distribution on \( \text{core}(\eta) \).

Adopt the latter specification to illustrate partial identification of \( \eta \). Consider an infinite sample for which the empirical frequency of \( B \) has the limit \( \ell > 0 \). The set of parameter values consistent with this sample is

\[
\Gamma_\ell = \{ \eta : \ell \leq \eta \} = [\ell, 1],
\]

which is the identified set given \( \ell \). For simplicity, let the prior \( \mu_0 \) have finite support. It follows as in Acemoglu et al. (2009, Lemma 1) that if \( \mu_0(\Gamma_\ell) > 0 \), then \( \lim_{n \to \infty} \mu_n(\Gamma_\ell) = 1 \) along the given sample. In fact,

\[
\lim_{n \to \infty} \mu_n(\eta) = \frac{\mu_0(\eta)}{\sum_{\eta' \in \Gamma_\ell} (\mu_0(\eta'))}, \quad \eta \in \Gamma_\ell.
\]

Thus there is asymptotic certainty only about the set \( \Gamma_\ell \). Note also that prior beliefs have a lasting effect and imply a posterior distribution within the identified set. Relative to the prior \( \mu_0 \), probability weight within \( \Gamma_\ell \) is shifted towards 'small' values of the parameter: for all \( \eta \) and \( \eta' \) with positive prior probability,

\[
\lim_{n \to \infty} \left( \frac{\mu_n(\eta)}{\mu_n(\eta')} \right) > \frac{\mu_0(\eta)}{\mu_0(\eta')} \quad \text{if} \quad \ell \leq \eta < \eta'.
\]

6. 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. Recall that in this binary case, each belief function \( \theta \) can be identified with a probability interval \( I_\theta = [\theta(B), \theta^*(B)] \) for \( B \), where \( \theta^*(B) = 1 - \theta(N) \).

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

\[
\max_{\alpha \in [0,1]} \int_{\text{Bel}(S)} \int_{\Omega} G (\Psi_n(\omega) - \alpha) d\theta^\infty d\mu(\theta), \quad (6.1)
\]
where $-G$ is a bounded strictly convex loss function that penalizes large differences between the predicted and realized frequencies $\alpha$ and $\Psi_n(\omega)$.

**Theorem 6.1.** There is a unique maximizer $\alpha_n$ in (6.1) and $\alpha_\infty \equiv \lim_{n \to \infty} \alpha_n$ exists. Moreover,

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

(6.2)

The limiting prediction $\alpha_\infty$ serves as an approximately optimal prediction for a sufficiently large number of experiments. Intuition for its characterization via (6.2) is derived from the LLN for i.i.d. belief functions (recall (4.3) and (4.4)). Fix $\theta$ and $\alpha$ and consider

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

(6.3)

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

$$
P\left(\{\lim \inf \Psi_n(\omega), \lim \sup \Psi_n(\omega)\} \subset [a, b]\right) = 0
$$

for some $P$ in $\text{core}(\theta^\infty)$. This suggests that, for large $n$, for the worst-case scenario in (6.3) it suffices to consider only samples that have empirical frequency equal to one of $\theta(B)$ and $\theta^*(B)$, as in (6.2).

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 so that the only relevant belief functions are of the form $\theta_\eta$ satisfying

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

Finally, assume certainty that $\eta$ is the true parameter value. Then (6.2) yields the closed-form solution

$$
\alpha_\infty = \frac{\eta}{2}.
$$
At the other extreme of predictions for a small number of markets, elementary calculations yield: \[\alpha_1 = \begin{cases} \eta & \eta \leq \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \leq \eta \end{cases}\] (6.4) and
\[\alpha_2 = \begin{cases} \eta & \eta \leq \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \leq \eta \leq \frac{1}{2} \\ \eta^2 & \frac{1}{2} \leq \eta \leq \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \leq \eta \end{cases}\] (6.5)

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 (2.4), the optimal prediction problem (when \(\mu(\theta) = 1\)) can be rewritten in the form

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

Then it follows from the minimax theorem that \(\alpha_n\) 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 \(\alpha_n\), that is, it solves

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

In brief, one can view \(\alpha_n\) as the best response to the scenario \(P^*\), and thus by identifying \(P^*\) we can understand the reasons for the choice of \(\alpha_n\). Apply the preceding to \(\alpha_2\) in (6.5). The corresponding worst-case measure \(P^*\) satisfies on \(\{B_1, N_1\} \times \{B_2, N_2\}\): \(\eta\)

\[
P^*(B_i) = \eta, \quad P^*(N_i) = 1 - \eta, \quad i = 1, 2, \text{ for all } \eta,
\]

and, if \(\eta > \frac{1}{2}\),

\[\text{Some details are provided in Appendix E.}\]
\[\text{See Appendix E for some details. Note that only the first two markets matter and thus we consider only the marginal of } P^* \text{ on } S_1 \times S_2.\]
\[ P^* (B_1, B_2) = \eta^2, \quad P^* (N_1, N_2) = 1 - \eta^2, \quad P^* (B_1, N_2) = P^* (N_1, B_2) = 0. \] (6.7)

Therefore, for \( \eta \) 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 one market then it is certain to be selected also in the other, and similarly for \( N \)). In contrast, 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 (6.6).

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

\[
\max_{\alpha \in [0,1]} \inf_{m \in \Delta([0,\eta])} \int_{\Delta(S)} \int_{\Omega} - (\alpha - \Psi_n (\omega))^2 dq^\infty dm (q). \quad (6.8)
\]

This decision-maker is also uncertain with which probability in \([0, \eta]\) 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.\(^\text{18}\) 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 (6.4) 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 \( \{B, N\} \) - the set of all distributions for which the probability of \( B \) is no greater than \( \eta \). However, they differ when predicting for two or more markets. In particular, \( \alpha^{RB}_2 \neq \alpha_2.\(^\text{19}\) We attribute this difference to the fact that only our agent is concerned about correlation between markets.\(^\text{20}\)

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

\[
\lim_{n \to \infty} \alpha^{RB}_n = \lim_{n \to \infty} \alpha_n = \alpha_\infty = \frac{\eta}{2}.
\]

\(^\text{18}\)Identify \( q \in \Delta(\{B, N\}) \) with the point \( q(B) \) in the unit interval.

\(^\text{19}\)One can compute that \( \alpha^{RB}_2 = \eta \) if \( \eta \leq \frac{1}{2} \) and \( \frac{1 + \eta}{2} \) otherwise.

\(^\text{20}\)A concern with heterogeneity does not seem relevant: if one generalizes the objective function in (6.8) by allowing nonidentical product measures \( q_1 \otimes q_1 \otimes \ldots \), the optimal prediction is not affected for any \( n \).
Thus the effects of ambiguity about correlation vanish in the limit.

[To be done: prediction problem where ambiguity about heterogeneity matters.]

7. Concluding Remarks

7.1. A comparison

Our objective was to generalize the exchangeable Bayesian model of inference and choice so as to accommodate the analyst’s concern that data may be correlated and heterogeneous in unknown ways. This was to be done while retaining as much as possible the attractive features of the Bayesian model, including in particular, its simple and intuitive axiomatic foundation. In order to facilitate judgement as to how well this has been accomplished, we list in the table below both sets of axioms.

<table>
<thead>
<tr>
<th>This Paper</th>
<th>Savage/de Finetti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief Function Utility, Weak Orthog. Indep.</td>
<td>SEU</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Symmetry</td>
</tr>
<tr>
<td>Consequentialism</td>
<td>Consequentialism</td>
</tr>
<tr>
<td>WEAK Dynamic Consistency</td>
<td>Dynamic Consistency</td>
</tr>
<tr>
<td>Commutativity</td>
<td>implied</td>
</tr>
</tbody>
</table>
| Functional form primitives \( (\mu_0, L) \) | \( \mu_0 \in \Delta (\Delta (S)) \) \\
\( \mu_0 \in \Delta (Bel (S)) \), \( L (\cdot | \theta) \in \Delta (S^\infty) \) exchangeable |

The gap between the two models 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.

7.2. 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 exoge-
nous 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 analyst believes that, for each \( i \), \( \epsilon_i = (\epsilon_{i1}, \epsilon_{i2}) \in \mathcal{E} \subset \mathbb{R}^2 \) is distributed according to \( m_\alpha \) and that \( \epsilon_i \)'s are i.i.d. The full set of parameters is \( \phi = (\alpha, \beta_1, \beta_2, \eta) \in \Phi \).

\[
\begin{array}{c|c|c}
\text{out} & \text{in} \\
\hline
0, 0 & 0, \beta_2 x_{i2} + \epsilon_{i2} \\
\beta_1 x_{i1} + \epsilon_{i1}, 0 & \beta_1 x_{i1} + \eta + \epsilon_{i1}, \beta_2 x_{i2} + \eta + \epsilon_{i2}
\end{array}
\]

Let \( Y = \{\text{out}, \text{in}\}^2 \) be the set of all pure strategy profiles in any single market. Given \( \phi \), 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 \( \Gamma_\phi : \mathcal{E} \rightarrow S \). Then the triple \((\mathcal{E}, m_\alpha, \Gamma_\phi)\) defines a belief function \( \theta_\phi \) on \( S \). The rest of the specification proceeds as before.

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 policy maker 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

\[ f^* (s_1, ..., s_i, ...) = u \left(s_1 \left(x^*\right), x^*\right) \text{ and } f^{**} (s_1, ..., s_i, ...) = u \left(s_1 \left(x^{**}\right), x^{**}\right). \]

Here \( s_1 \left(x^*\right) \) 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 \( s_1 \left(x^{**}\right) \).

### 7.3. 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.” As noted earlier, stochastic independence is more complicated in the nonadditive probability (or multiple-priors) framework and there is more than one way to form independent products (see Ghirardato (1997)). Accordingly, the representation in our previous paper admits various product rules. In contrast, in (1.3) the rule for forming the i.i.d. product $\theta^\infty$ 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 4.1) a choice-theoretic rationale for any particular i.i.d. product rule. The value added herein lies also in the connection drawn between ambiguity and partial identification, as in games with multiple equilibria. That connection is original to this paper.

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 (6.8)), in that they deal with ambiguity about parameters but exclude ambiguity about how experiments are related; for example, they cannot exhibit the rankings (3.2) and (3.3). As a result these models seem orthogonal to the central issues raised by multiple equilibria in entry games. Moreover, they model 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.

---

21 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.
A. Appendix: Belief Functions

The following notation is used throughout the appendices. For any compact metric space \( \Omega \), \( \mathcal{K}(\Omega) \) is the space of compact subsets endowed with the Hausdorff metric; \( \Delta(\Omega) \) is the space of Borel probability measures on \( \Omega \) endowed with the weak convergence topology; and \( \text{Bel}(\Omega) \) 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 \( \Omega \), where the integral is in the sense of Choquet. All three spaces are compact metric. They are endowed with the corresponding Borel \( \sigma \)-algebras.

For any metric space \( X \), its \( \sigma \)-algebra is denoted \( \mathcal{X} \).

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 \( \Omega \), which can be any compact metric space. The set of acts \( \mathcal{F} \) is defined as in the text.

A belief function is most commonly defined as a set function \( \nu : \Sigma_\Omega \to [0,1] \) satisfying:

\begin{align*}
\text{Bel.1} & \quad \nu(\emptyset) = 0 \text{ and } \nu(\Omega) = 1 \\
\text{Bel.2} & \quad \nu(A) \leq \nu(B) \text{ for all Borel sets } A \subset B \\
\text{Bel.3} & \quad \nu(B_n) \downarrow \nu(B) \text{ for all sequences of Borel sets } B_n \downarrow B \\
\text{Bel.4} & \quad \nu(G) = \sup \{ \nu(K) : K \subset G, K \text{ compact} \}, \text{ for all open } G \\
\text{Bel.5} & \quad \nu \text{ is totally monotone (or } \infty\text{-monotone): for all Borel sets } B_1, \ldots, B_n, \\

\nu \left( \bigcup_{j=1}^n B_j \right) & \geq \sum_{\emptyset \neq J \subset \{1, \ldots, n\}} (-1)^{|J|+1} \nu \left( \bigcap_{j \in J} B_j \right)
\end{align*}

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 \( \nu \) is convex (super-modular, 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.\footnote{Throughout, given any Borel probability measure, we identify it with its unique extension to the $\sigma$-algebra of universally measurable sets. Below $P$ is short-hand for $\int f dP$.}

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

$$\nu(A) = m_\nu(\{K \in \mathcal{K}(\Omega) : K \subseteq A\})$$

for every $A \in \Sigma_{\Omega}$. \hspace{1cm} (A.1)

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

$$U_\nu(f) = \int_{\Omega} f d\nu = \int_{\mathcal{K}(\Omega)} \left( \inf_{P \in \Delta(K)} P : f \right) dm_\nu(K)$$

$$= \int_{\mathcal{K}(\Omega)} \left( \inf_{x \in K} f(x) \right) dm_\nu(K).$$

We use frequently below the implication that every belief function (at least as defined by Bel.1-Bel.5) on a space $\Omega$ 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 Section 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 $\hat{\Omega} = \mathcal{K}(\Omega)$, $\Gamma$ maps any $K$ (a point in $\mathcal{K}(\Omega)$) into $K$ (a subset of $\Omega$) and $m = m_\nu$. Conversely, let $\nu$ be defined via the triple $\left(\hat{\Omega}, m, \Gamma^{\nu}\right)$ and (2.1)-(2.3). View $\Gamma$ as a function from $\hat{\Omega}$ to $\mathcal{K}(\Omega)$. Then $\Gamma$ is measurable (Aliprantis and Border (2006, Thm. 18.10)) and induces the measure $m^{\nu} = m \circ \Gamma^{-1}$ on $\mathcal{K}(\Omega)$. Then Choquet’s Theorem implies that $\nu(\cdot) = m \circ \Gamma^{-1}(\{K : K \subseteq \cdot\})$ satisfies Bel.1-Bel.5 and $m^{\nu} = m_\nu$.

Now let $\Omega = S^\infty$. Let $\theta \in Bel(S)$ be generated by $(\hat{\Omega}, m, \Gamma)$. We defined $\theta^\infty$ to be the belief function on $\Omega$ represented by $(\hat{\Omega}, m^\infty, \Gamma^\infty)$, where: $\hat{\Omega} = \left(\hat{S}\right)^\infty$, $m^\infty$ is the ordinary i.i.d. product of the probability measure $m$, and $\Gamma^\infty$ is the correspondence $\Gamma^\infty : \hat{\Omega} \rightarrow \hat{\Omega} = S^\infty$ given by (2.7). Choquet’s theorem gives an alternative characterization of the product that we use frequently below. In particular, it implies that the product $\theta^\infty$ is independent of the particular representation $(\hat{\Omega}, m, \Gamma)$ for $\theta$. 

22Throughout, given any Borel probability measure, we identify it with its unique extension to the $\sigma$-algebra of universally measurable sets. Below $P \cdot f$ is short-hand for $\int f dP$. 

31
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 $(m_\theta)^\infty$).

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 $\mathcal{K}(\Omega)$. The proof of the lemma is omitted.

B. Appendix: Proof of Theorem 4.1

First we prove the measurability required to show that the integrals in (4.1) and (4.2) are well-defined. (Recall that the Borel probability measure $\mu$ has a unique extension to the class of all universally measurable subsets.)

Lemma B.1. Both $\theta \mapsto V_{\theta^\infty}(f)$ and $\theta \mapsto \theta^\infty(A)$ are universally measurable for any $f \in \mathcal{F}$ and $A \in \Sigma_\Omega$.

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

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

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

Step 2. $P \mapsto \int \hat{f} dP$ from $\Delta([\mathcal{K}(S)]^\infty)$ to $\mathbb{R}$ is analytically measurable for any bounded analytically measurable function $\hat{f}$ on $[\mathcal{K}(S)]^\infty$: If $\hat{f}$ is simple (has a finite number of values), then $P \mapsto \int \hat{f} dP$ 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

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

is coanalytic by Phillipe et al. (1999, p. 772), and hence analytically measurable.
Steps 1, 2 and 3 complete the proof.

For Theorem 4.1, we show (iii)⇒(ii)⇒(i)⇒(iii). For any \( \nu \in \text{Bel} (\Omega) \), denote by \( \zeta (\nu) \) the measure \( m_\nu \) on \( K (\Omega) \) provided by the Choquet theorem. (Similarly if \( \theta \in \text{Bel} (S) \), then \( \zeta (\theta) \in \Delta (K (S)) \).) We use (A.2) repeatedly without reference.

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

\[
\{ K \in K : K \subset A \}_{A \in \Sigma}.
\]

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

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

for each \( A \in \Sigma \). This is equivalent to

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

which is true given (iii).

By a standard argument using the Lebesgue Dominated Convergence Theorem,

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

for all \( \Sigma' \)-measurable \( f : K (\Omega) \rightarrow [0, 1] \). Since \( K \mapsto \inf_{\omega \in K} f (\omega) \) is \( \Sigma' \)-measurable by (B.1),

\[
U_\nu (f) = \int_{K(\Omega)} \inf_{\omega \in K} f (\omega) \, dm_\nu (K) = \int_{\text{Bel}(S)} \left( \int_{K(\Omega)} \inf_{\omega \in K} f (\omega) \, d\zeta (\theta^\infty) \right) \, d\mu (\theta)
\]

\[
= \int_{\text{Bel}(S)} V_{\theta^\infty} (f) \, d\mu (\theta).
\]

(ii)⇒(i): It is enough to show that \( V_{\theta^\infty} \) satisfies Symmetry and WOI. Let \( m = \zeta (\theta^\infty) \). My Lemma A.2, \( m \) is an i.i.d. measure on \( [K (S)]^\infty \), hence symmetric.
Therefore,

\[
V_{\theta}(\pi f) = \int_{K(\Omega)} \inf_{\omega \in K} \pi f(\omega) \, dm(K) = \int_{K(\Omega)} \inf_{\omega \in K} f(\pi \omega) \, dm(K)
\]

\[
= \int_{K(\Omega)} \inf_{\pi \omega \in \pi K} f(\pi \omega) \, dm(K) = \int_{K(\Omega)} \inf_{\omega \in K} f(\omega) \, d(\pi m)(K)
\]

\[
= \int_{K(\Omega)} \inf_{\omega \in K} f(\omega) \, dm(K) = V_{\theta}(f).
\]

Show (3.4) to prove WOI. For simplicity, let \(f \in F_1\) and \(g \in F_2\). The general case is similar. For \(0 < \alpha \leq 1\),

\[
V_{\theta}(\alpha f + (1 - \alpha) g)
\]

\[
= \int_{K(\Omega)} \inf_{\omega \in K} [\alpha f(\omega) + (1 - \alpha) g(\omega)] \, dm(K)
\]

\[
= \int_{[K(S)]^\infty} \inf_{s_1 \in K_1, s_2 \in K_2} [\alpha f(s_1) + (1 - \alpha) g(s_2)] \, dm(K_1, K_2, ...)
\]

\[
= \int_{[K(S)]^\infty} \alpha \left[ \inf_{s_1 \in K_1} f(s_1) \right] + (1 - \alpha) \left[ \inf_{s_2 \in K_2} (1 - \alpha) g(s_2) \right] \, dm(K_1, K_2, ...)
\]

\[
= \alpha \int_{[K(S)]^\infty} \left[ \inf_{s_1 \in K_1} f(s_1) \right] \, dm(K_1, K_2, ...)
\]

\[
+ (1 - \alpha) \int_{[K(S)]^\infty} \left[ \inf_{s_2 \in K_2} g(s_2) \right] \, dm(K_1, K_2, ...)
\]

\[
= \alpha V_{\theta}(f) + (1 - \alpha) V_{\theta}(g).
\]

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

(i)⇒(iii): For \(C \subset K(\Omega)\), let \(\pi C = \{\pi K \in K(\Omega) : K \in C\}\), and for \(m \in \Delta(K(\Omega))\), define \(\pi m \in \Delta(K(\Omega))\) by \(\pi m(C) = m(\pi C)\) for each Borel measurable \(C \subset K(\Omega)\).

**Lemma B.2.** For any \(m \in \Delta(K(\Omega))\), \(m = \pi m\) for all \(\pi\) if and only if \(m = \zeta(\nu)\) for some symmetric belief function \(\nu\) on \(\Omega\).
Proof. If \( m = \zeta (\nu) \), then \( \nu (K) = m (\{K' \in \mathcal{K} (\Omega) : K' \subset K\}) \), and

\[
\nu (\pi K) = m (\{K' \in \mathcal{K} (\Omega) : K' \subset \pi K\}) = m (\{\pi K' \in \mathcal{K} (\Omega) : \pi K' \subset \pi K\}) = m (\{\{K' \in \mathcal{K} (\Omega) : K' \subset K\}\}.
\]

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

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

Proof. For any \( \omega \in \Omega \) and disjoint sets \( I, J \subset \mathbb{N}, \omega_I \) denotes the projection of \( \omega \) onto \( S^n \), and we write \( \omega = (\omega_I, \omega_J, \omega_{1...I}) \). When \( I = \{i\} \), we write \( \omega_i \), rather than \( \omega_{\{i\}} \), to denote the \( i \)-th component of \( \omega \).

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

\[
\exists \omega^* \in K, \text{ such that } \omega^*_I = \omega^1_I \text{ and } \omega^*_J = \omega^2_J. \quad (B.2)
\]

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 \( J \),

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

a.s.-\( m \)[\( K \)]: This is where WOI enters - by (3.4) it implies that

\[
U_\nu (\frac{1}{2} f + \frac{1}{2} g) = \frac{1}{2} U_\nu (f) + \frac{1}{2} U_\nu (g).
\]

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

\[
\int_{\mathcal{K} (\Omega)} \inf_{\omega \in K} \left[ \frac{1}{2} f (\omega) + \frac{1}{2} g (\omega) \right] \, d m (K) = \frac{1}{2} \int_{\mathcal{K} (\Omega)} \inf_{\omega \in K} f (\omega) \, d m (K) + \frac{1}{2} \int_{\mathcal{K} (\Omega)} \inf_{\omega \in K} g (\omega) \, d m (K).
\]

The assertion follows from

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

**Step 2.** \( m \left( \bigcap_{(f,g) \in G} \mathcal{B}_{f,g} \right) = 1 \): 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 \( G \) is also separable. Let \( \{(f_n, g_n)\} \) be a countable dense subset of \( G \). By Step 1,

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

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

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

\[
\min_{\omega \in K} \left[ \frac{1}{2} f(\omega) + \frac{1}{2} g(\omega) \right] = \lim_{i} \min_{\omega \in K} \left[ \frac{1}{2} f_i(\omega) + \frac{1}{2} g_i(\omega) \right] = \lim_{i} \left[ \frac{1}{2} \min_{\omega \in K} f_i(\omega) + \frac{1}{2} \min_{\omega \in K} g_i(\omega) \right] = \frac{1}{2} \min_{\omega \in K} f(\omega) + \frac{1}{2} \min_{\omega \in K} g(\omega).
\]

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

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

\[
A_i = \left\{ \omega : \sum_{t \in I} 2^{-t} d(\omega_t, \omega_t^1) \geq \frac{1}{i} \right\} \quad \text{and} \quad B_i = \left\{ \omega : \sum_{t \in J} 2^{-t} d(\omega_t, \omega_t^2) \geq \frac{1}{i} \right\},
\]

36
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 \( i \),

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

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 \( K \in B_{f_i,g_i} \),

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

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

\[
\sum_{t \in I} 2^{-t} d(\hat{\omega}^t, \omega_I) + \sum_{t \in J} 2^{-t} d(\hat{\omega}^t, \omega_J) < \frac{2}{i}.
\]

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, \( 1 \geq m(\mathcal{A}) \geq m\left( \bigcap_{(f,g) \in \mathcal{G}} B_{f,g} \right) = 1 \).

**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 \( \omega^1, \omega^2, \ldots \in K \). It suffices to show that

\[
\omega^* = (\omega_1^1, \omega_2^2, \ldots, \omega_n^n, \ldots) \in K.
\]  

(B.4)

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

Finally, we prove (i)⇒(ii). Let \( \nu \) be a belief function on \( \Omega \) and suppose that \( U_\nu \) 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

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

37
Here each $\ell$ lies in $\Delta(K(S))$ and $\ell^\infty$ is the i.i.d. product measure on $[K(S)]^\infty$. Extend each measure $\ell^\infty$ to $\Sigma_{K(\Omega)}$ and write

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

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

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

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

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

$$\nu(A) = m(\{K \in K(\Omega) : K \subset A\})$$

$$= \int_{\Delta(K(S))} \ell^\infty(\{K \in K(\Omega) : K \subset A\}) d\mu(\ell)$$

$$= \int_{\Delta(K(S))} \ell^\infty(\{K \in K(\Omega) : K \subset A\}) d\mu \circ \zeta^{-1}(\ell)$$

$$= \int_{Bel(S)} [\zeta(\theta)]^\infty (\{K \in K(\Omega) : K \subset A\}) d\mu(\theta)$$

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

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

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

C. Appendix: Proof of Corollary 4.2

Because $\theta^\infty(A) = \theta(A)$ for $A \in \Sigma_S$, $\theta \mapsto \theta(A)$ is universally measurable by Lemma B.1. Hence, every set of the form

$$\{\theta \in Bel(S) : [\theta(A), 1 - \theta(S \setminus A)] \subset [a, b]\}$$
is universally measurable and the statement of the Corollary is well-defined.

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

Lemma C.1. Let \( K \in [K(S)]^\infty \), \( K = K_1 \times K_2 \times ... \), 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) \( \liminf_n \hat{\Psi}_n (A) (K) > \alpha \).

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

(ii) \( \Rightarrow \) (i): If \( s_i \in K_i, I(K_i \subset A) \leq I(s_i \in A) \). Thus, if \( s_i \in K_i \) for \( i = 1, ... \), then,
\[ \liminf_n \Psi_n (A) (\omega) \geq \liminf_n \hat{\Psi}_n (A) (K) > \alpha. \]

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

Proof. Fix \( A \in \Sigma_S \). Then,
\[ \theta^\infty (\{ \omega : \theta (A) < \liminf_n \Psi_n (A) (\omega) \}) \]
\[ = [\zeta(\theta)]^\infty (\{ K \in [K(S)]^\infty : K \subset \{ \omega : \theta (A) < \liminf_n \Psi_n (A) (\omega) \} \}) \]
\[ = [\zeta(\theta)]^\infty (\{ K \in [K(S)]^\infty : \liminf_n \hat{\Psi}_n (A) (K) > \theta (A) \}) \] (by Lemma C.1).

By the classical LLN, \( \hat{\Psi}_n (A) (K) \) converges to 
\( \zeta(\theta) (\{ K_1 \in K(S) : K_1 \subset A \}) = \theta (A) \) almost surely-\( [\zeta(\theta)]^\infty \), which implies (i).

The proof of (ii) is similar.

Return to the Corollary. By the LLN in Maccheroni and Marinacci (2005), Lemma C.2 and the monotonicity of belief functions,
\[ \theta^\infty (\{ \omega : [\liminf_n \Psi_n (A) (\omega), \limsup_n \Psi_n (A) (\omega)] \subset [a, b] \}) = 1 \]
\[ \iff [\theta (A), 1 - \theta (S \setminus A)] \subset [a, b] \]
and
\[ \theta^\infty (\{\omega : [\liminf_n \Psi_n (A) (\omega), \limsup_n \Psi_n (A) (\omega)] \subset [a, b]\}) = 0 \]
\[ \iff [\theta (A), 1 - \theta (S \setminus A)] \text{ is not a subset of } [a, b]. \]

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

\[ \nu \left( \bigcap_{j=1}^J \{\omega : [\liminf_n \Psi_n (A_j) (\omega), \limsup_n \Psi_n (A_j) (\omega)] \subset [a_j, b_j]\} \right) \]
\[ = \int_{Bel(S)} \theta^\infty \left( \bigcap_{j=1}^J \{\omega : [\liminf_n \Psi_n (A_j) (\omega), \limsup_n \Psi_n (A_j) (\omega)] \subset [a_j, b_j]\} \right) d\mu (\theta) \]
\[ = \nu \left( \bigcap_{j=1}^J \{\theta : [\theta (A_j), 1 - \theta (S \setminus A_j)] \subset [a_j, b_j]\} \right). \]

**Proposition C.3.** If \( \mu \) and \( \mu' \), probability measures on \( Bel(S) \), coincide on all sets of the form

\[ \{\theta \in Bel(S) : \theta (A_1) \geq a_1, \ldots, \theta (A_J) \geq a_J\}, \]

where \( A_j, a_j \) and \( J \) vary over \( \Sigma_S \), \([0, 1]\) and the positive integers respectively, then \( \mu = \mu' \).

**Proof.** We can identify \( \mu' \) and \( \mu \) with measures on \( \Delta (K (S)) \). Modulo this identification, we are given that \( \mu' \) and \( \mu \) agree on the collection of all subsets of \( \Delta (K (S)) \) of the form

\[ \bigcap_{j=1}^J \{\ell \in \Delta (K (S)) : \ell (\{K \in K (S) : K \subset A_j\}) \geq a_j\}, \]

for all \( J > 0 \), \( A_j \in \Sigma_S \) and \( a_j \in [0, 1] \). They necessarily agree also on the generated \( \sigma \)-algebra, denoted \( \Sigma^* \). Therefore, it suffices to show that

\[ \Sigma_{\Delta (K(S))} \subset \Sigma^*. \]
Step 1. \( \ell \mapsto \ell (C) \) is \( \Sigma^* \)-measurable for measurable \( C \in \Sigma_{\mathcal{K}(S)} \): Let \( \mathcal{C} \) be the collection of measurable subsets \( C \) of \( \mathcal{K}(S) \) such that \( \ell \mapsto \ell (C) \) is \( \Sigma^* \)-measurable. Every set of the form \( \{ K' \in \mathcal{K}(S) : K' \subset K \} \) for \( K \in \mathcal{K}(S) \) lies in \( \mathcal{C} \). Since the collection \( \{ K' \in \mathcal{K}(S) : K' \subset K \} \) generates \( \Sigma_{\mathcal{K}(S)} \), it is enough to show that \( \mathcal{C} \) is a \( \sigma \)-algebra: (i) \( C \in \mathcal{C} \) implies \( \mathcal{K}(S) \setminus C \in \mathcal{C} \); (ii) if each \( C_j \in \mathcal{C} \), then \( \ell \mapsto \ell \left( \bigcup_{j=1}^{\infty} C_j \right) \) is \( \Sigma^* \)-measurable because it equals the pointwise limit of \( \ell \mapsto \ell \left( \bigcup_{j=1}^{n} C_j \right) \) - hence \( \bigcup_{j=1}^{\infty} C_j \in \mathcal{C} \).

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

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

\[ \text{D. Appendix: Proof of Theorem 5.1} \]

Consider the following modified statement: Suppose all preferences can be represented as in Theorem 4.1. Then the axioms WDC and Consequentialism are satisfied if and only if there exists a likelihood function \( L : \text{Bel}(S) \rightarrow \Delta(S^\infty) \) such that posteriors are generated by applying Bayes’ rule to \( \mu_0 \) and \( L \). This is a minor modification of Theorem 6.1 in Epstein and Seo (2010) 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 assertions dealing with Commutativity. Its necessity given (5.1) is clear. Thus assume Commutativity (and the modified result just discussed, which gives a not necessarily exchangeable likelihood \( L \)) and 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 \( \pi \). By uniqueness of the representing measure in Theorem 4.1,

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

Let

\[ \mathcal{L} (s^n) \equiv \int L (s^n \mid \theta) d\mu_0. \]
Then
\[ U_0(f) = \sum_{s^n} L(s^n) U_n(f \mid s^n) \]
\[ = \sum_{k=0}^{n} \sum_{s^n \in S^{n,k}} L(s^n) U_n(f \mid s^n) \]
\[ = \sum_{k=0}^{n} \left( \sum_{s^n \in S^{n,k}} L(s^n) \right) U_n(f \mid s^n), \]

where \( S^{n,k} \) is the set of all samples \( s^n \in S^n \) with \( k \) occurrences of \( B \). Define \( L^*_n \in \Delta(S^n) \) by
\[ L^*_n(s^n) = \frac{1}{|S^{n,k}|} \left( \sum_{s^n \in S^{n,k}} L(s^n) \right) \text{ if } s^n \in S^{n,k}. \]

By the Kolomogorov Extension Theorem, there exists \( L^* \in \Delta(S^\infty) \) that coincides with \( L^*_n \) on \( S^n \) for every \( n \). Therefore, it is exchangeable and satisfies, for every \( n \),
\[ U_0(\cdot) = \sum_{s^n} L^*(s^n) U_n(\cdot \mid s^n) \text{ on } \mathcal{F}. \] \( \text{(D.1)} \)

The latter equation leads to the desired likelihood function. Take
\[ L^*(s^n \mid \theta) = L^*(s^n) (d\mu(\theta \mid s^n) / d\mu_0(\theta)). \]

By the uniqueness of representing measures in Theorem 4.1, (D.1) implies
\[ \mu_0(\cdot) = \sum_{s^n} L^*(s^n) \mu(\cdot \mid s^n), \]
and thus \( \Sigma_{s^n} L^*(s^n \mid \theta) = 1 \) for all \( \theta \). Further,
\[ L^*(s^n \mid \theta) = L^*(s^n) (d\mu(\theta \mid s^n) / d\mu_0(\theta)) \]
\[ = L^*(\pi s^n) (d\mu(\theta \mid \pi s^n) / d\mu_0(\theta)) \]
\[ = L^*(\pi s^n \mid \theta). \]

It is easily verified that posteriors are generated by Bayesian updating using \( \mu_0 \) and \( L^* \) (proceed as in the proof of Theorem 6.1 of our earlier paper).
E. Appendix: Prediction

This appendix deals with binary experiments, \( S = \{B, N\} \). Each belief function \( \theta \) 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 E.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

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

that is, there exists a constant \( K \) such that

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

**Proof of Theorem 6.1:** Step 1: Show that

\[
\lim_{n \to \infty} \int \int G(\alpha -\Psi_n(\omega)) d\theta^\infty d\mu_0(\theta)
\]


\[
= \int \min\{G(\alpha -\theta(B)),G(\alpha -\theta^*(B))\} d\mu_0(\theta).
\]

By the CLT, for each \( \theta \),

\[
\int G(\alpha -\Psi_n(\omega)) d\theta^\infty = E[\min\{G(\alpha -\bar{X}_{1n}),G(\alpha -\bar{X}_{2n})\}] + O\left(\frac{1}{\sqrt{n}}\right),
\]
where $\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}$ and $\bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$, and each $(X_{1i}, X_{2i})$ is normally distributed (i.i.d. across $i$’s) with mean $(\theta(H), 1 - \theta(T))$ and variance 

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

By the classical strong LLN, $(\bar{X}_1, \bar{X}_2)$ converges to $(\theta(B), \theta^*(B))$ a.s. with respect to the i.i.d. product of the above normal. Then, by the continuous mapping theorem, $\min \{G(\alpha - \bar{X}_1) : G(\alpha - \bar{X}_2)\}$ converges to $\min \{G(\alpha - \theta(B)) : G(\alpha - \theta^*(B))\}$ a.s. and thus in distribution. Therefore,

$$E \left[ \min \{G(\alpha - \bar{X}_1), G(\alpha - \bar{X}_2)\} \right] \to \min \{G(\alpha - \theta(B)), G(\alpha - \theta^*(B))\},$$

(E.1)

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

Step 2: Show that

$$\lim_{n \to \infty} \arg\max_{\alpha \in [0,1]} \int \int G(\alpha - \Psi_n(\omega)) d\theta^\infty d\mu(\theta) = \arg\max_{\alpha \in [0,1]} \int \int G(\alpha - \Psi_n(\omega)) d\theta^\infty d\mu(\theta).$$

The set $\{1, 2, ..., \infty\}$ is compact when endowed with the topology generated by singletons $\{n\}$ and sets of the form $\{n, ..., \infty\}$. Define $F : [0,1] \times \{1, 2, ..., \infty\} \to [-1,0]$ 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 [details to be provided]. There is a unique solution $\alpha_n$ for $\max_{\alpha \in [0,1]} \int \int G(\alpha - \Psi_n(\omega)) d\theta^\infty d\mu(\theta)$: Obviously the maximum exists. Uniqueness follows from the strict concavity of $\alpha \mapsto \int G(\alpha - \Psi_n(\omega)) d\theta^\infty$ for each $\theta$. Application of the Maximum Theorem completes the proof of this step.

Step 3: Complete the proof. From Steps 1 and 2,
\[
\alpha_\infty \equiv \lim_{n \to \infty} \arg\max_\alpha \int \int G(\alpha - \Psi_n(\omega))d\theta^\infty d\mu(\theta)
= \arg\max_\alpha \lim_{n \to \infty} \int \int G(\alpha - \Psi_n(\omega))d\theta^\infty d\mu(\theta)
= \arg\max_\alpha \min \{G(\alpha - \theta(B)), G(\alpha - \theta^*(B))\} d\mu(\theta). \]

To be provided: Details regarding calculations in Section 6.

References


[23] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique priors, 


[26] E. Hewitt and L.J. Savage, Symmetric measures on Cartesian products, 


[35] F. Maccheroni and M. Marinacci, A strong law of large numbers for capacities, 


