AMBIGUITY WITH REPEATED EXPERIMENTS*

Larry G. Epstein Kyoungwon Seo

May 11, 2012

Abstract

The paper generalizes de Finetti’s exchangeable Bayesian model of beliefs and preference for a setting with repeated experiments. The model is axiomatic and describes the behavior of a decision-maker who has a theory or model of her environment, but who is sophisticated enough to realize both that she has little information about the factors that are common across experiments and that her theory of how experiments are related is incomplete. A motivating example is policy choice in the context of a discrete entry game with multiple equilibria such as has been studied in the literature on partial identification.

Keywords: ambiguity, repeated experiments, exchangeable measures, incomplete theory, partial identification, robust statistics, maxmin expected utility, multiple priors, multiple equilibria, entry games

*Boston University, lepstein@bu.edu and Department of Managerial Economics and Decision Sciences, Northwestern University, k-seo@kellogg.northwestern.edu. We are grateful to Chew Soo Hong, Hiro Kaido, Peter Klibanoff and Minjae Song for helpful conversations. This research was supported by the National Science Foundation (awards SES-0917740 and SES-0918248). Epstein acknowledges the hospitality of the Einaudi Institute for Economics and Finance where substantial revisions of the paper were completed. First version: August 1, 2010.
1. Introduction

1.1. Motivation and objectives

Though subjective expected utility (SEU) theory continues to be the dominant model of choice under uncertainty, it has been assailed at both normative and descriptive levels as being unable to accommodate a role for ambiguity, or confidence in beliefs, such as illustrated by the Ellsberg Paradox. Motivated by this critique, Gilboa and Schmeidler (1989) axiomatize the maxmin expected utility (MEU) model, a generalization of SEU. The popularity of SEU is due in part to the elegant and appealing axiomatizations provided by Savage and by Anscombe and Aumann. However, its prominence in statistical decision-making and in learning models is due also in large part to de Finetti (1937) who considered settings with repeated experiments and introduced added structure that is intuitive for such settings. De Finetti showed that the simple property of exchangeability characterizes beliefs for which outcomes of experiments are i.i.d. conditional on an unknown parameter. Our objective is to characterize a corresponding specialization of MEU that is intuitive and useful given uncertainty represented by repeated experiments. In the rest of this introduction we elaborate on the nature of this specialization and the reasons for pursuing it.

Consider a setting with a countable number of experiments, each of which yields an outcome in the set \( S \) (technical details are suppressed until later). Thus \( \Omega = S^\infty \) is the set of all possible sample paths. A probability measure \( P \) on \( \Omega \) is exchangeable if the probability of a sequence of outcomes does not depend on the order in which they are realized. De Finetti shows that exchangeability is equivalent to the following representation: There exists a (necessarily unique) probability measure \( \mu \) on \( \Delta(S) \), the set of probability laws on \( S \), such that

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

(1.1)

where, for any probability measure \( \ell \) on \( S \), \( \ell^\infty \) denotes the corresponding i.i.d. product measure on \( \Omega \).\(^1\) Kreps (1988) refers to this result as “the fundamental theorem of (most) statistics,” because of the justification it provides for the analyst to view samples as being independent and identically distributed with unknown

\(^1\)Though the de Finetti theorem can be viewed as a result in probability theory alone, it is typically understood in economics as describing beliefs in the subjective expected utility model of choice. That is how we view it in this paper.
distribution function. In addition, it is widely viewed as providing foundations for modeling beliefs about an unknown “parameter,” here the true probability law \( \ell \) on \( S \).

But the representation (1.1) is suggestive of features that might be viewed as overly restrictive, in both statistical analysis and decision-making more broadly. First, uncertainty about the probability law \( \ell \) is expressed via the single prior \( \mu \), and thus, as is familiar from the literature inspired by the Ellsberg Paradox, the model precludes ambiguity about the true parameter. Second, uncertainty is limited to the true law describing any single experiment - the individual is certain that experiments are both identical (the same law \( \ell \) applies to every one) and independent. Both features have been criticized (implicitly) as presuming too much confidence on the part of the statistical decision-maker. Robust statistics (Huber (1981)) is motivated by the nonrobustness of standard procedures related to the Ellsberg critique of modeling beliefs by a single prior. For another illustration of why the exchangeable Bayesian model might be viewed as too restrictive, think of a cross-sectional empirical model where experiments correspond to the regression errors and where these errors are not well understood. For example, the decision-maker may be concerned that some relevant variables have been omitted from the regressions, though she cannot be specific about which variables and about the consequences of their omission - some omitted variables may influence specific experiments in the same direction and others in opposite directions, while other omitted variables may have different effects. The decision-maker simply does not understand the regression errors in her model well enough to be more specific. But she is sophisticated enough to realize this and wishes to take it into account when making decisions. That this is an important issue in applied economics is emphasized in the opening sentence of Ibragimov and Müller (2010): “Empirical analyses in economics often face the difficulty that the data are correlated and heterogeneous in some unknown fashion.” A related concrete example is described in the next subsection.

Our objective is to generalize the de Finetti model, viewed as normative model, to simultaneously accommodate both ambiguity about parameters and a concern with possible heterogeneity (and correlation) of experiments. Our model is axiomatic (see particularly Theorem 6.1). Thus it provides behavioral meaning and foundations for decision-making guided by the following perception: The outcomes of experiments depend on both common factors, or “parameters”, about which prior beliefs may be imprecise, and on factors that vary across experiments in some unknown or poorly understood fashion and that render experiments hetero-
geneous (and possibly correlated). This description would seem to apply to many choice settings where the decision-maker has a theory or model of her environment, but where she is sophisticated enough to realize both that she has limited information about the true parameter and that her theory is incomplete, by which we mean that even knowledge of the true parameter would not be enough to permit probabilistic predictions about future experiments. Such incompleteness is illustrated in the next subsection and in Section 5, and its meaning in terms of behavior is described also in the discussion of our axioms.

We have emphasized the presence of ambiguity about both parameters (common factors) and about how experiments differ from and are related to one another (heterogeneity and correlation). This begs the question whether a behavioral distinction is possible between these two intuitive but informal notions. Every preference in our model exhibits aversion to both kinds of ambiguity. A distinction is established through intuitive definitions of indifference to each kind of ambiguity separately. Theorem 4.1 describes the corresponding representations. They can be understood via reference to the de Finetti representation (1.1), which features a single prior and also a single (uncertain) likelihood that describes each experiment. Roughly speaking, our general model features both multiple priors and multiple likelihoods describing each experiment. It is shown that the special case of a single likelihood (and multiple priors) corresponds to indifference to ambiguity about heterogeneity and correlation, and that the special case of a single prior (and multiple likelihoods) corresponds to indifference to ambiguity about parameters.

Our previous paper Epstein and Seo (2010), henceforth ES, also studies ambiguity given repeated experiments. It characterizes each of the two special representations just described. However, it does not include a general model where there is aversion to both sources of ambiguity. The present model provides a unifying framework that also permits a sharper behavioral distinction between parameters and heterogeneity as sources of ambiguity than is provided in ES (see Section 4).

Another related paper is Epstein and Schneider (2007). They are the first to emphasize the distinction between common and variable factors driving experiments; our running example of a sequence of Ellsberg urns is adapted from one of their examples. One difference from this paper is that their model is designed for a setting where experiments are ordered in time, preference is constructed so as to be recursive, and updating is the central concern. In contrast, our model is designed for cross-sectional experiments, where symmetry, analogous to de Finetti’s ex-
changeability, is more natural, and we do not consider updating. Another major
difference is that they describe functional forms and provide informal justification,
partly through applications, while here the focus is on axiomatic foundations.

We adopt Gilboa and Schmeidler’s MEU as the basis for our model of preference
given repeated experiments. There exist other models of ambiguity averse
preferences that might be adopted as a basis, for example, variational utility
(Maccheroni et al. (2006)) which generalizes MEU. However, generality is not our
goal. Rather, consistent with standard methodological practice, we seek a minimal
departure from SEU that permits accommodating the behavior of interest. The
maxmin model seems to us to work well in this regard. Of course, a generalization
to variational utility could be justified by a suitable behavioral critique of MEU,
particularly one that is germane to the setting of repeated experiments.

The paper proceeds as follows. Next in this introduction we describe two
motivating examples. Then, after taking care of some formal preliminaries, we
describe the functional form for utility that we propose. Its axiomatic foundations
are described later in Section 6. Beforehand, we characterize the two special cases
noted above (Section 4) and then we describe how the utility function can be
used to address the two examples (Section 5). Concluding remarks are offered in
Section 7.

1.2. Two examples

We provide two running examples. The first deals with Ellsberg-style urns. The
second concerns the more concrete problem of policy choice in the context of a
complete information entry game with multiple equilibria such as considered in
Tamer (2003) and Ciliberto and Tamer (2009), for example. We will revisit them
in Section 5 after laying out our model to show how it performs in both cases.

Example 1: Urns

As a running example, consider a sequence of Ellsberg urns. You are told that
each contains 100 balls that are either red or blue, thus $S = \{R, B\}$. You may
also be given additional information, symmetric across urns, but it does not pin
down either the precise composition of each urn or the relationship between urns.
In particular, the information is consistent with the urns differing in composition.
One ball will be drawn from each urn, which defines “experiment.” You must

\footnote{See Epstein and Seo (2011a) for clarification of the relation between the
time-series and cross-sectional versions of the model.}
choose between bets on the outcomes of the sequence of draws. The ranking of
bets depends on how the experiments are perceived. One conceivable view of the
urns is that they all have identical compositions, with the unknown proportion \( \theta \) of
red balls. In the de Finetti model, beliefs would be represented by a single prior
over possible values of the parameter \( \theta \). More generally, \( \theta \) may be ambiguous.
Regardless, it is part of a complete theory of the experiments - knowledge of \( \theta \)
would imply a unique (probabilistic) prediction of outcomes. Alternatively, one
might behave as if the urns are unrelated. In this case, the only common factor
is that the proportion of red balls lies in \([0, 1]\) for each urn, which interval thus
constitutes the (here known) parameter. The parameter is trivial in the sense that
knowing it does not help to predict outcomes, reflecting the perception of extreme
heterogeneity. More generally, we strive to capture also intermediate perceptions.
For example, one might perceive that the fraction \( \lambda \) of the 100 balls is selected
by a single administrator and then placed in each urn, while the other \((1 - \lambda) 100\)
varies across urns in a way that is not understood (as in the extreme perception
above). If \( \theta \) denotes the proportion of red in the common group of balls, then the
probability of drawing red from any urn lies between \( \lambda \theta \) and \( \lambda \theta + (1 - \lambda) \). Thus the
unknown parameter can be thought of as the probability interval for red given by
\[ I = [\lambda \theta, \lambda \theta + (1 - \lambda)] \]. In general, there will be ambiguity both about the degree
\( \lambda \) of commonality across urns and about the color composition of the common
group. Thus ambiguity about the correct interval \( I \) is to be expected. Note that
knowledge of \( I \) would be useful for prediction, though it does not yield a unique
probability. In addition, since no information is provided about the relationship
between urns, one would expect ambiguity also about the color composition of
the variable component consisting of \((1 - \lambda) 100\) balls, and about how that color
composition varies across urns.

Section 5 describes behavior that is intuitive in this setting and demonstrates
how our model accommodates it. Because our model is the only one in the liter-
ature that can do so, this serves as behavioral motivation for the model.

**Example 2: Multiple equilibria**

There are \( I \) markets. In the \( i^{th} \) market two players (who differ across markets)
play the entry game with payoffs (in utils) given by the following matrix:

\[
\begin{array}{ccc|cc}
 & \text{out} & \text{in} \\
\text{out} & 0, 0 & 0, 1 - \epsilon_{2i} \\
\text{in} & 2 - \epsilon_{1i}, 0 & \theta_1 - \epsilon_{1i}, \theta_2 - \epsilon_{2i} \\
\end{array}
\]
The parameter $\theta = (\theta_1, \theta_2) \in \Theta$ does not vary across markets, but the random factor $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})$ does vary. Let each $\varepsilon_i$ take on values in the set $\mathcal{E}$. The set of possible outcomes in each market is $S = \{0, 1\} \times \{0, 1\}$, where $(1, 0)$ indicates that player 1 chooses in and player 2 chooses out, and so on. Assume that both $\theta$ and $\varepsilon_i$ are known to the players and that they play a mixed strategy Nash equilibrium.

There is also a policy choice to be made before observing the equilibrium outcomes in any markets. For concreteness adapt the set up in Ciliberto and Tamer (2009) by taking players to be airlines and markets to be trips between pairs of airports. The decision maker is the government who is contemplating constructing one small or a large airport. An important consideration is that one airline is large (player 1) and the other is small (player 2) in every market. Thus, were the airport to serve only one market, then a small airport would be preferable if serving only the small airline and the large airport would be superior if the large airline were to enter, either alone or with the smaller one. In this case, construction of a small airport could be thought of as a bet on the outcome $(0, 1)$ and choice of a large airport would amount to a bet on $\{(1, 0), (1, 1)\}$. More generally, the value of the airport is tied to all markets that begin or end at that airport, and thus a construction plan is an act, that is, a mapping from $S^I$ to the space of payoffs, where $I$ consists of markets that begin or end at the constructed airport.\footnote{The example assumes that the size of the new airport does not affect the payoff of the players. In Section 5, we demonstrate that our modeling framework can accommodate also policies that affect payoffs of the airlines.}

The policy maker knows the structure of the game but she does not know the values of $\theta$ or the $\varepsilon_i$’s. She views the latter as distributed i.i.d. across markets according to the measure $m \in \Delta(\mathcal{E})$. She has some information about $\theta$, perhaps from past experience in other markets, and forms a set of priors about its values. Another important aspect of the environment is that there may be multiple Nash equilibria in any market. Furthermore, the policy maker does not understand how selection occurs and how it may differ across markets. Thus her theory of the environment is incomplete - probabilistic predictions of outcomes would be impossible even if she knew the parameter $\theta$. This incompleteness leaves the door open to heterogeneity and correlation across markets that she cannot describe but that she may suspect is important. How should she choose between available policies?

In her ignorance of the selection mechanism and consequent reluctance to theorize about it, our policy maker is similar to the statistical decision makers
described in a strand of the partial identification literature; see, for example, Tamer (2003, 2010), Ciliberto and Tamer (2009) and the references therein. There also the individual wishes to make decisions (albeit statistical decisions as opposed to the economic decisions modeled here) and wishes to base them only on what she understands, while refusing to ignore or deny her limited understanding. This leads to statistical procedures that distinguish between two kinds of uncertainty, one that can be eliminated with enough data (think of \( \theta \) here) and the other (corresponding to the selection mechanism) that persists because one cannot learn from data in the absence of a theory, and that leads to only a (nonsingleton) set of parameters being identified. Alternatively, in terminology that we prefer, the identified set can be thought of as the parameter, and partial identification is then reflected in the fact that even given knowledge of that parameter only a nonsingleton set of likelihoods is implied.

In Section 5, we show how the policy choice problem can be modeled using our “exchangeable” version of MEU. (As noted there, a number of generalizations of the choice problem can be accommodated.) Note that the partial identification literature focuses on identification (and inference) and has less to say about how to make economic decisions based on the identified set; see, however, Manski (2011) and the references therein, and Kasy (2011). The connection between partial identification and axiomatic models of ambiguity averse preference was made by Epstein and Seo (2011b). However, their model precludes ambiguity about the parameter \( \theta \) and accommodates only pure strategy Nash equilibria.

2. Formal Preliminaries

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

\[
\Omega = S^\infty = S_1 \times S_2 \times \ldots
\]

Denote by \( \Sigma \) the product \( \sigma \)-algebra on \( \Omega \). Probability measures on \( (\Omega, \Sigma) \) are understood to be countably additive unless specified otherwise.

An act is a \( \Sigma \)-measurable function from \( \Omega \) into \([0, 1]\). The set of all acts is \( \mathcal{F} \). For any subset \( I \) of \( \mathbb{N} \), \( \Sigma_I \) denotes the product \( \sigma \)-algebra on \( \prod_{i \in I} S_i \), also identified with a \( \sigma \)-algebra on \( \Omega \), and \( \mathcal{F}_I \) denotes the set of all acts that are \( \Sigma_I \)-measurable. (When \( I = \{i\} \), we write \( \Sigma_i \) and \( f \in \mathcal{F}_i \).

Such acts will be said to depend only
on experiments in $I$. Particularly important are acts that depend on finitely many experiments, that is, acts in
$$\mathcal{F}_{\text{fin}} = \cup_{I \text{ finite}} \mathcal{F}_I.$$ Refer to such acts as \textit{finitely-based}. Say that $f^*$ and $f$ are (mutually) \textit{orthogonal}, written $f^* \perp f$, if $f^* \in \mathcal{F}_{I^*}$ and $f \in \mathcal{F}_I$ for some disjoint $I^*$ and $I$. Any $c$ in $[0, 1]$ is identified with the act that is constant at $c$. Note that every constant act is orthogonal to every act ($c \perp f$ since we can take $I^*$ to be the empty set).

Denote by $\Pi$ the set of finite permutations of $\mathbb{N}$; all permutations appearing below should be understood to be finite unless specified otherwise. For any $\pi$ in $\Pi$ and any probability measure $P$ on $(S^\infty, \Sigma)$, define $\pi P$ to be the unique probability measure on $S^\infty$ satisfying (for all rectangles)
$$(\pi P) (A_1 \times A_2 \times \ldots) = P(A_{\pi^{-1}(1)} \times A_{\pi^{-1}(2)} \times \ldots).$$

Given an act $f$, define the permuted act $\pi f$ by $(\pi f) (s_1, \ldots, s_t, \ldots) = f (s_{\pi(1)}, \ldots, s_{\pi(t)}, \ldots)$. Abbreviate $\int fdP$ by $P f$, or $P (f)$. Then, for all $P, f$ and $\pi$,
$$(\pi P) f = P (\pi f).$$

The probability measure $P$ is \textit{exchangeable} if $\pi P = P$ for all $\pi$.

Preference, denoted $\succeq$, is defined on $\mathcal{F}$. We assume that $\succeq$ is a MEU preference, by which we mean that it has a representation of the following form:\footnote{For any compact metric space $X$, $\Delta (X)$ denotes the set of countably additive Borel probability measures on $X$, endowed with the weak-convergence topology induced by continuous functions. $\mathcal{K} (X)$ denotes the space of compact subsets of $X$, endowed with the Hausdorff metric topology, which renders it compact metric. When $X$ is a linear topological space, $\mathcal{K}_c (X)$ denotes the subspace of compact and convex subsets of $X$.} There exists a (weak-convergence) compact and convex set $\mathcal{P} \subset \Delta (\Omega)$, such that
$$W (f) = \inf_{\mathcal{P}} Pf = \inf_{\mathcal{P}} \int fdP, \ f \in \mathcal{F}. \quad (2.1)$$

Behavioral foundations for this variant of the Gilboa-Schmeidler model are provided in ES. An important feature of this model is that the ranking of finitely-based acts is enough to determine preference on all acts. (The proof follows from ES (Thm. 7.1) and Epstein and Wang (1995, Thm. D.2).)
Lemma 2.1. Let $\succsim'$ and $\succsim$ each have utility functions as in (2.1) with sets $\mathcal{P}'$ and $\mathcal{P}$ respectively. If $\succsim'$ and $\succsim$ coincide on $\mathcal{F}_{\text{fin}}$, then $\mathcal{P}' = \mathcal{P}$ and thus $\succsim'$ and $\succsim$ coincide on $\mathcal{F}$.

The reader will have noticed that acts are taken to be real-valued and that they enter linearly into the utility calculation in (2.1). This should be understood as a normalization justified in the familiar way by an Anscombe-Aumann formulation. A consequence is that the utility $W(f)$ is scaled in probability units - it satisfies

$$f \sim (1, W(f); 0, 1 - W(f)).$$

(2.2)

Thus $f$ is indifferent to the lottery giving 1 (the best outcome) with probability $W(f)$ and 0 (the worst outcome) with probability $1 - W(f)$. In particular, for any outcome $c$, the above indifference allows us to interpret $c = W(c)$ as the lottery $(1, c; 0, 1 - c)$, or as the associated probability of winning.

The interpretation of outcomes as probabilities helps to interpret the construct of product acts. Given any two acts $f^*$ and $f$, then $f^* \cdot f$ denotes the pointwise product, that is, the act given by

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

If $f^*$ is the bet on drawing red from the first urn and $f$ is the bet on drawing blue from the second urn, both with prizes 1 and 0, then $f^* \cdot f$ is the bet on the intersection, namely on the first draw being red and the second being blue. In general, the outcome produced by $f$ in state $\omega$ can be viewed as a bet on heads in the toss of a coin with objective probability for heads equal to $f(\omega)$. Similarly, in state $\omega$ the product act $f^* \cdot f$ corresponds to a bet on heads where the objective probability of heads is $f^*(\omega) f(\omega)$, or equivalently to a bet on successive heads in independent tosses of the two coins associated with $f^*$ and $f$.

A final preliminary matter concerns terminology. The de Finetti representation (1.1) involves two probability measures - the measure $\mu$ over the unknown parameter and the implied probability measure $\mathcal{P}$ over the payoff-relevant state space $\Omega$. To distinguish between them, we adopt the common practice of referring to the former as the prior and to the latter as the predictive prior (or distribution). Similarly, for MEU functions, refer to $\mathcal{P}$ as the set of predictive priors, and to the set $M$ of measures that is the counterpart of the single $\mu$ (see (3.5)) as the set of priors.
3. A Functional Form for Utility

First consider the benchmark SEU utility function specialized according to de Finetti. If the predictive prior is exchangeable and has the form in (1.1), then the corresponding SEU function has the form

$$W_{SEU}(f) = \int_{\Delta(S)} V_{\ell}(f) \, d\mu(\ell). \tag{3.1}$$

Here $f$ denotes the generic act (or generalized bet) over $\Omega$ (with outcomes denominated in utils), and $V_{\ell}(\cdot)$ is the expected utility function with i.i.d. measure $\ell^\infty$,

$$V_{\ell}(f) = \int_\Omega f(\omega) \, d\ell^\infty(\omega). \tag{3.2}$$

To describe our generalization of (3.1), we define a broader notion of i.i.d. Say that the MEU utility function $V$ is a *product utility function* if

$$V(f \cdot g) = V(f) V(g) \text{ for all orthogonal } f, g \in \mathcal{F}_{fin}. \tag{3.3}$$

It satisfies Symmetry if

$$V(f) = V(\pi f) \text{ for all } \pi \in \Pi \text{ and all } f \in \mathcal{F}_{fin}. \tag{3.4}$$

Refer to $V$ as an *IID utility function* if it is a product utility function satisfying Symmetry; refer to the corresponding set of priors $\mathcal{P}$ as an *IID set of priors*. The set of IID utility functions is denoted $\mathcal{V}$.\(^5\)

Following Epstein and Schneider (2003), the acronym IID stands for “independently and indistinguishably (as opposed to identically) distributed.”\(^6\) Given SEU preferences, an IID utility function is an SEU function where the predictive prior is i.i.d. However, stochastic independence is more complicated in the MEU (or nonadditive probability) framework in that there is more than one way to form independent products; see Hendon et al. (1996) and Ghirardato (1997), and also Section 5 below for examples. The property (3.3) expresses a weak form of

---

\(^5\)Any set of predictive priors $\mathcal{P}$ lies in $\mathcal{K}^c(\Delta(\Omega))$, the space of compact and convex subsets of $\Delta(\Omega)$; the Hausdorff metric topology renders it compact metric. Each $\mathcal{P} \in \mathcal{K}^c(\Delta(\Omega))$ corresponds to a unique MEU preference, or equivalently, to a unique MEU utility function. Under this identification, $\mathcal{V}$ inherits a compact metric topology.

\(^6\)We continue to use the lower case acronym i.i.d. when referring to single measures, with the usual meaning of “independently and identically distributed.”
stochastic independence between experiments that is satisfied by all notions of independent products we are aware of in the literature. Therefore, by not restricting IID utility functions further to have particular functional forms, our model provides a unifying framework for a range of particular specifications (see Section 5 for two such examples). Note also that (3.3) leaves scope for modeling ambiguity about correlation (see (5.12) below).

Now we can define our general model of utility $W : \mathcal{F} \to \mathbb{R}$, which has the form:

$$W(f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V(f) \, d\mu(V), \ f \in \mathcal{F},$$

for some (weak-convergence) compact and convex set of priors $\mathcal{M} \subset \Delta(\mathcal{V})$. It is shown in Appendix A.1 that the function $W$ so-defined can be expressed in the MEU form (2.1) for some compact and convex set of predictive priors $\mathcal{P} \subset \Delta(\Omega)$.

The set $\mathcal{M}$ represents prior beliefs about the unknown IID utility function $V$. Were $V$ known with certainty, then by (3.4) it would rank bets on any single experiment; in that sense one can view $V$ as common to all experiments and hence as a parameter. However, $V$ is unknown in general. This interpretation is clear in the special case (3.1) corresponding to the de Finetti representation, which is obtained if $\mathcal{M} = \{\mu\}$, where $\mu$ has support on the subset of $\mathcal{V}$ consisting of SEU utility functions. Then each $V$ corresponds to a likelihood function which is the unknown parameter of interest. Further interpretation and illustration follow in the sequel.

Say that the set of priors $\mathcal{M}$ represents the preference $\succeq$ on $\mathcal{F}$ if $W$ in (3.5) is a utility function for $\succeq$. It is easily seen that the representing set is typically not unique since one could add to $\mathcal{M}$ measures that would not affect the infimum in (3.5) regardless of the act $f$. For example, if $\mu \in \mathcal{M}$ and if $\mu'$ is a probability measure on $\mathcal{V}$ satisfying

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

then the infimum in (3.5) is unchanged if we use $\mathcal{M} \cup \{\mu'\}$ instead of $\mathcal{M}$. (This is obvious if we restrict attention to finitely-based acts in (3.5), but it is valid more generally for all acts for the reasons underlying Lemma 2.1.) Less obvious is that

\footnote{Technical details ensuring that the right-hand side of (3.5) is well-defined are provided in Appendix A in the context of the proof of Theorem 6.1.}

\footnote{More precisely, to see the connection with (3.1), identify $\mu$ with a measure on $\Delta(S)$.}
the above describes completely the nature of nonuniqueness in the representation, as we now make precise.

For any two probability measures on $\mathcal{V}$, write $\mu' \triangleright \mu$ if (3.6) is satisfied. Given any $\mathcal{M} \subseteq \Delta(\mathcal{V})$, define its $\triangleright$-hull $\widehat{\mathcal{M}}$ by

$$\widehat{\mathcal{M}} = \{ \mu' \in \Delta(\mathcal{V}) : \mu' \triangleright \mu \text{ for some } \mu \in \mathcal{M} \}.$$ 

In general, $\widehat{\mathcal{M}} \supseteq \mathcal{M}$. Say that $\mathcal{M}$ is comprehensive if $\widehat{\mathcal{M}} = \mathcal{M}$.

**Theorem 3.1.** Every preference $\succeq$ on $\mathcal{F}$ admits representation by at most one comprehensive, compact and convex set of priors $\mathcal{M}$.

Finally, compare this uniqueness result to what is known from Gilboa and Schmeidler about uniqueness in the MEU model in an abstract setting. Their uniqueness result, adapted to the variant of their model defined by (2.1), is that the set of predictive priors representing any preference is unique - see Lemma 2.1. In contrast, in our setting of repeated experiments, where the representation (3.5) applies, the corollary deals with uniqueness of the set of (non-predictive) priors representing preference. The difference is easily illustrated: Let $V_1$ and $V_2$ be two IID utility functions with sets of predictive priors $Q_1$ and $Q_2$, subsets of $\Delta(\Omega)$. Let $Q_2 \subseteq Q_1$, so that $V_2(\cdot) \succeq V_1(\cdot)$ on $\mathcal{F}$. Suppose finally that $\mu$ assigns probability 1 to $V_1$ and that $\mu'$ has support equal to $\{V_1, V_2\}$. Then

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

Therefore, the sets of priors $\{\mu, \mu'\}$ and $\{\mu\}$ represent the same preference or utility function $W$. Indeed, consistent with Lemma 2.1, they generate the same set of predictive priors for $W$ equal to $Q_1$.

4. Special Cases and a Behavioral Distinction

In the introduction we emphasized the distinction, germane to settings with repeated experiments, between ambiguity about parameters (or common factors) and ambiguity about how experiments differ and are related (heterogeneity and correlation). Here we show how the functional form (3.5) captures this distinction

---

9That there exist distinct measures satisfying this relation is clear from (5.5).

10For brevity, we often refer to ambiguity about heterogeneity, though the intention is typically to ‘heterogeneity and correlation.’
by examining special cases. We also give behavioral meaning to the distinction through alternative axioms for preference.

Refer to the single prior model when preference admits representation by a singleton set of priors $\mathcal{M} = \{\mu\}$, that is,

$$W(f) = \int V(f) \, d\mu(V), \quad f \in \mathcal{F}. \quad (4.1)$$

Refer to the single likelihood model when preference admits representation by a set of priors $\mathcal{M}$ all of whose measures have support in $\mathcal{V}_{SEU}$, the subset of $\mathcal{V}$ consisting of SEU utility functions. Every $V$ in $\mathcal{V}_{SEU}$ has the form in (3.2) - $V = V_\ell$ for some $\ell$ in $\Delta(S)$ - and $\ell \mapsto V_\ell$ defines an isomorphism between $\mathcal{V}_{SEU}$ and $\Delta(S)$. Therefore, with a slight abuse of notation, the representation can be written in the form

$$W(f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}_{SEU}} V(f) \, d\mu(V) = \inf_{\mu \in \mathcal{M}} \int_{\Delta(S)} V_\ell(f) \, d\mu(\ell), \quad f \in \mathcal{F}. \quad (4.2)$$

The functional forms suggest the following interpretations: the single likelihood model reflects ambiguity about the correct likelihood, (which is the unknown ‘parameter’), but also certainty that experiments are i.i.d. for some $\ell$. In contrast, the single prior model suggests ambiguity only about heterogeneity - each $V$ corresponds to a set $Q$ of probability measures for any single experiment, where the set is common to all experiments but where different measures in $Q$ could apply to different experiments, thus admitting heterogeneity. The behavioral definitions and axiomatic characterizations to follow provide behavioral support for these interpretations and also demonstrate the unifying power of the general model.

Following Gilboa and Schmeidler, mixtures of acts play an important role throughout. For any acts $f$ and $g$, and for any weight $\alpha$ in the unit interval, the mixture $\alpha f + (1 - \alpha) g$ is defined by

$$(\alpha f + (1 - \alpha) g)(\omega) = \alpha f(\omega) + (1 - \alpha) g(\omega), \quad \omega \in \Omega.$$ 

Gilboa and Schmeidler identify a value for such randomization because it can smooth outcomes across ambiguous states. If $f$ and $f'$ are indifferent, say that they hedge one another if $\alpha f + (1 - \alpha) f' \succ f$ for all $\alpha$ in $(0, 1)$.\footnote{The condition “for some $\alpha$ in $(0, 1)$” yields an equivalent definition (this follows from Lemma A.1).} Otherwise, say that $f$ and $f'$ do not hedge one another.
Say that preference is averse to ambiguity about heterogeneity if:\(^{12}\) For all \(h, h' \in F_{\{1, \ldots, n\}}\), and \(\alpha \in [0, 1]\),
\[
(\alpha h + (1 - \alpha) h') \succsim (\alpha h + (1 - \alpha) T^n h').
\]
(4.3)

Interpret the condition in the context of the urns example. (Denote by \(R_n\) the bet that yields the prize 1 if red is drawn from the \(n^{th}\) urn and 0 otherwise; interpret \(R_1B_2, \alpha R_2 + (1 - \alpha) R_1\) and so on in the obvious way.) Then (4.3) requires, for example, that
\[
\frac{1}{2} R_1 + \frac{1}{2} B_1 \succsim \frac{1}{2} R_1 + \frac{1}{2} B_2,
\]
which has clear intuition: The mixture on the left eliminates all ambiguity, (indeed, its payoff is \(\frac{1}{2}\) with certainty), while the mixture on the right moderates only ambiguity about factors that are common across experiments, that is, about parameters. If the individual cares about possible heterogeneity, then \(\frac{1}{2} R_1 + \frac{1}{2} B_2\) leaves her worse off. Similarly for the comparison in (4.3), where the mixture of orthogonal acts \(h\) and \(T^n h\) limits the gains relative to those from mixing \(h\) and \(h'\). If there is indifference in (4.3) for all acts indicated there, say that preference is indifferent to ambiguity about heterogeneity.

Refer to aversion to ambiguity about parameters if: For all \(n, h, h' \in F_{\{1, \ldots, n\}}\), and \(\alpha \in [0, 1]\),
\[
\mathcal{W}(\alpha h + (1 - \alpha) T^n h') \geq \alpha \mathcal{W}(h) + (1 - \alpha) \mathcal{W}(T^n h').
\]
(4.4)

Note that because \(\mathcal{W}\) gives the probability equivalents of acts (recall (2.2)), this is a restriction on preference. In the urns example, (4.4) requires that\(^{13}\)
\[
\mathcal{W}(\frac{1}{2} R_1 + \frac{1}{2} B_2) \geq \frac{1}{2} \mathcal{W}(R_1) + \frac{1}{2} \mathcal{W}(B_2).
\]
The mixture on the left mixes bets on distinct urns and thus does not hedge ambiguity about idiosyncratic variations across urns. However, it may be valuable if there is a perception that there exists a common factor, (for example, if the fraction \(\lambda\) of balls in every urn is selected by a single administrator and thus is identical in composition), and if the composition of these 100\(\lambda\) balls is unknown. Then mixing between bets on red and blue is valuable, as in the classic Ellsberg

\(^{12}\)\(T\) denotes the shift operator, \((Tf)(s_1, s_2, s_3, ...) = f(s_2, s_3, ...)\); \(T^n\) denotes the \(n\)-fold replication of \(T\).

\(^{13}\)When \(R_1 \sim B_1\), and hence also \(R_1 \sim B_2\), (4.4) can be expressed equivalently in terms of preference as \(\frac{1}{2} R_1 + \frac{1}{2} B_2 \succeq R_1\).
experiment, even where the bets are on the draws from different urns. Such mixing
hedges ambiguity about the proportion of red in the common 100\lambda balls, which
is the ambiguous parameter in this case. The intuition is similar for the general
case of (4.4). If there is indifference in the latter for all acts indicated there, say
that preference is \textit{indifferent to ambiguity about parameters}.

\textbf{Theorem 4.1.} \textit{Let the preference} \(\succeq\) \textit{have a utility function of the form (3.5). Then: (i) \(\succeq\) is averse to ambiguity about both heterogeneity and parameters. (ii) \(\succeq\) is indifferent to ambiguity about parameters if and only if it conforms to the single prior model. Moreover, in that case, the representing prior} \(\mu\) \textit{is unique. (iii) \(\succeq\) is indifferent to ambiguity about heterogeneity if and only if it conforms to the single likelihood model. Moreover, in that case, the representing (compact and convex) set of priors} \(\mathcal{M}\) \textit{is unique. (iv) \(\succeq\) is indifferent to ambiguity about both heterogeneity and parameters if and only if it conforms to the de Finetti model.}

Statement (i) is readily verified. For example, consider ambiguity about heterogeneity. Then

\[ W(\alpha h + (1 - \alpha) T^n h') = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V(\alpha h + (1 - \alpha) T^n h') d\mu(V) \]

\[ = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} [\alpha V(h) + (1 - \alpha) V(T^n h') d\mu(V) \]

\[ = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} [\alpha V(h) + (1 - \alpha) V(h')] d\mu(V) \]

\[ \leq \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} [V(\alpha h + (1 - \alpha) h') d\mu(V) = W(\alpha h + (1 - \alpha) h'), \]

where use has been made of the fact that each \(V\) is MEU, satisfies Symmetry (from which it follows that \(V(T^n h') = V(h')\)), and is linear when mixing orthogonal acts (see Appendix A.1).

For (iv), note that indifference to both kinds of ambiguity implies

\[ W(\alpha h + (1 - \alpha) h') = W(\alpha h + (1 - \alpha) T^n h') \]

\[ = \alpha W(h) + (1 - \alpha) W(T^n h') = \alpha W(h) + (1 - \alpha) W(h'), \]

which gives the Independence axiom on \(\mathcal{F}_{\text{fin}}\). See Appendix D for proofs of (ii) and (iii).
Remark 1. When combined with the axiomatization of (3.5) in Theorem 6.1, (call these Axioms*), the preceding provides an axiomatization of both the single likelihood and the single prior models. These are axiomatized also in ES. An advantage of the present characterization is that it isolates better the behavioral difference between the two models, which here takes the form of indifference to ambiguity about heterogeneity (IH) versus indifference to ambiguity about parameters (IP). In contrast, the characterization in ES can be interpreted as inextricably lumping Axioms* together with each of IH and IP and characterizing the two models respectively by [Axioms*+IH] and [Axioms*+IP]. The implicit and inseparable common component Axioms* blurs the distinction between the models.

5. Modeling the Two Examples

We revisit the two examples described in Section 1.2 and show how they would be treated in our model. In the context of the urns example we describe behavior that intuitively reflects ambiguity about parameters and also about heterogeneity. We show how our model accommodates this behavior and indicate where specializations of the model fail to do so. In particular, the behavioral case against the de Finetti-SEU model is reinforced. The second example demonstrates that our model can be applied naturally to concrete settings that have been of interest in the empirical IO literature.

Because the functional form (3.5) relies heavily on IID utility functions, we first describe two such functions, borrowing from our papers ES and (2011b). The fact that there exist many different IID utility functions reflects the fact, noted earlier, that stochastic independence is multi-faceted in the multiple-priors framework, and hence that there is more than one way to form an independent product from a given set of priors over $S$. In other words, in contrast to the Bayesian setting, there are many utility functions satisfying (3.3), and hence the “stochastic independence” embodied in it, that also agree on the ranking of acts over any single experiment.

For the first IID utility function, fix a (closed) set $\mathcal{L}$ of probability measures on $S$, thought of as the set of probability laws applying to any single experiment. The same set $\mathcal{L}$ applies to each experiment but different experiments could be
described by distinct measures in \( \mathcal{L} \). Thus let \( \mathcal{L}^\infty \subset \Delta (\Omega) \) be given by\(^{14}\)

\[
\mathcal{L}^\infty \equiv \{ \otimes_{i \in \mathbb{N}} \ell_i : \ell_i \in \mathcal{L} \text{ for every } i \} ,
\]

and define utility \( V_{WF} \) by\(^{15}\)

\[
V_{WF} (f) = \inf_{P \in \mathcal{L}^\infty} Pf, \quad f \in \mathcal{F} .
\]

Since \( \mathcal{L}^\infty \) consists exclusively of product measures, (3.3) is obvious; so is Symmetry. Therefore, \( V_{WF} \) is an IID utility function. The subscript WF indicates that we have adapted this specification from Walley and Fine (1982).

The second IID utility function also begins with a set \( \mathcal{L} \) of probability measures on \( S \), but the latter is restricted to be the core of a belief function, that is,

\[
\mathcal{L} = \text{core} (\nu) \equiv \{ p \in \Delta (S) : p (\cdot) \geq \nu (\cdot) \} ,
\]

for some belief function \( \nu \).\(^{16}\) A central fact about belief functions is that each belief function \( \nu \) can be represented equivalently by a probability measure \( m \) on \( \mathcal{K}(S) \) such that

\[
\text{core} (\nu) = \{ \sum_{K \in \mathcal{K}(S)} m (K) p_K : p_K \in \Delta (K) \} \equiv \sum_{K \in \mathcal{K}(S)} m (K) \Delta (K) .
\]

The IID product of \( \mathcal{L} \) is constructed by using the i.i.d. product measure \( m^\infty \in \Delta (\mathcal{(K(S))}^\infty) \) and applying a corresponding formula to define the set \( \otimes_H \mathcal{L} \subset \Delta (\Omega) \) by\(^{17}\)

\[
\otimes_H \mathcal{L} = \int_{(\mathcal{K}(S))^\infty} \Delta (K_1 \times K_2 \times ...) \, dm^\infty (K_1, K_2, ... ) .
\]

The subscript \( H \) indicates that we have adapted this specification from Hendon et al. (1996). The corresponding utility function \( V_H \) is

\[
V_H (f) = \inf_{P \in \otimes_H \mathcal{L}} Pf
\]

\[
= \int_{(\mathcal{K}(S))^\infty} \left( \inf_{P \in \Delta (K_1 \times K_2 \times ...)} P \cdot f \right) \, dm^\infty (K_1, K_2, ... ) .
\]

\(^{14}\)\( \otimes_{i \in \mathbb{N}} \ell_i \) denotes the unique countably additive product measure with marginals \( \ell_i \).

\(^{15}\)\( \mathcal{L}^\infty \) is not convex, but the identical utility is defined if we use instead its closed convex hull.

\(^{16}\)For references to literature on the theory and applications of belief functions see our paper (2010b).

\(^{17}\)The integral is defined as an Aumann integral. See Appendix A.1.
Then $V_H$ is an IID utility function. (Symmetry is clear and the product rule (3.3) is readily verified.) This is so even though, unlike $L_1$, $H$ does not consist exclusively of product measures.

Because they share the set $L$ of predictive priors for a single experiment, $V_{WF}$ and $V_H$ agree on $F_n$, for every $n$. However, they differ on acts that depend on more than a single experiment, as illustrated and exploited below. Moreover, because $L^\infty \subset \otimes H L$, it follows that

$$V_{WF} (\cdot) \geq V_H (\cdot) \text{ on } F.$$  \hspace{1cm} (5.5)

There exist other IID utility functions - see Couso et al. (1999) and the references therein for examples where there are only finitely many experiments, some of which can be extended to the present setting of infinitely many experiments. However, the above two specifications suffice to illustrate our model.

The Urns Example Revisited

Turn to the urns example, so that $S = \{R, B\}$. The binary case permits a simplified description. Each set $L$ can be identified with a unique interval $I = [p_m, p_M] \subset [0, 1]$, thought of as an interval of probabilities for red. Therefore, each such interval corresponds also to an IID utility function $V_{WF}$ via (5.2). Similarly for $V_H$. For every interval $I$ there exists a belief function $\nu$ that generates it: $\nu (R) = p_m$, $\nu (B) = 1 - p_M$, and $\nu (\{R, B\}) = 1$. In particular, for a binary state space, the belief function assumption is unrestrictive and any belief function can be thought of simply as a probability interval. The corresponding measure $m$ on the set of subsets of $S$ is given by $m (R) = p_m$, $m (B) = 1 - p_M$ and $m (\{R, B\}) = p_M - p_m$, the length of the interval. Consequently, $I$ defines, via $m$ and (5.4), a unique IID utility function $V_H$.

For concreteness, suppose only three intervals are relevant - $I^0 = [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$, $I = [\frac{7}{16}, \frac{11}{16}]$ and its reflection about $\frac{1}{2}$, $I' = [\frac{5}{16}, \frac{9}{16}]$. Roughly, every urn is seen as being consistent with a color composition that is unbiased (having an equal number of balls of each color), but the degree and nature of the ambiguity vary with the interval: it is either small and symmetric in the two colors (in the case of $I^0$), or admits a possibly large bias towards red (in the case of $I$), or towards blue (in the case of $I'$). Denote by $V_{WF}^0, V_{HW}^0, V_{WF}, V_{WF}'$ and $V_H'$ the corresponding IID utility functions. For our purposes, it suffices to restrict attention to the subset $\mathcal{V}^0 = \{V_{WF}^0, V_H, V_{HW}^0\}$ of these IID utility functions. In other words, let preference be represented by the utility function (3.5), where all priors in $\mathcal{M}$
assign probability 1 to $V_0$. The utility function can be written in the form

$$W(f) = \min_{\mu \in \mathcal{M}} \left\{ \mu \left( V_{0W}^0 \right) V_{0W}^0(f) + \mu (V_H) V_H(f) + \mu (V'_H) V'_H(f) \right\}. \quad (5.6)$$

Because information about the urns gives no reason for distinguishing between red and blue, it is compelling that, for every urn the individual be indifferent between betting on drawing red or drawing blue. Evidently, $W(R_i) = W(B_i)$ given our construction. We describe a number of other less immediate properties of preference that are intuitive given our story and that are satisfied under suitable assumptions on the set of priors $\mathcal{M}$.

A simple sufficient (but far from necessary) assumption is that $\mathcal{M} = \{ \mu \in \Delta (V^0) : \mu (V_{0W}^0), \mu (V_H), \mu (V'_H) \geq \epsilon \}$. \quad (5.7)

where $0 < \epsilon < \frac{1}{3}$.

Consider bets on the color of the ball drawn from the first urn. Because these are ambiguous, the individual values randomization in that

$$\frac{1}{2} R_1 + \frac{1}{2} B_1 \succ R_1 \sim B_1. \quad (5.8)$$

This is a version of Ellsberg’s two-urn example - the mixture gives the outcome $\frac{1}{2}$ with certainty while the bets on red and blue, though indifferent to one another, are ambiguous and hence strictly less attractive than the sure bet. (Each of $V_{0W}^0, V_H$ and $V'_H$ ranks the mixture as strictly preferable and thus so does $W$.)

Another intuitive ranking is

$$\frac{1}{2} R_1 + \frac{1}{2} B_2 \succ R_1. \quad (5.9)$$

The intuition is similar to that for (5.8) in that bets on opposite colors moderate ambiguity about the composition of each urn. The difference here is that the bets concern draws from different urns. As described above, the hedging value is still positive if, as one would expect in general, the individual perceives a common element to the compositions of different urns. We have referred to this as ambiguity about parameters; one can think of the probability intervals, $I^0$, $I$ or $I'$, as the unknown parameter that is common across urns. (In terms of the functional form, suppose that $\delta < \frac{3}{16}$. Then $V'_H(R)$ is smaller than $V_{0W}^0 (R)$ and $V_H (R)$. Therefore, the bad scenarios for the bet on red are those attaching high probability to

\footnotetext[18]{Detailed verification of (5.8)-(5.12) below is left to the reader. We provide only some informal supporting arguments.}
I’, or \( V_H’ \). Similarly, the bad scenarios for the bet on blue are those attaching high probability to \( I \), or \( V_H \). Randomizing between these bets is beneficial because it smooths across these worst-case scenarios.)

For a third ranking, consider

\[
\frac{1}{2} R_1 + \frac{1}{2} B_1 \succ \frac{1}{2} R_1 + \frac{1}{2} B_2. 
\]

This is an instance of heterogeneity ambiguity aversion defined in (4.3) and interpreted there. Though the same probability interval, \( I^0, I \) or \( I’ \), describes every urn, the individual may be concerned that even given an interval, different probabilities in it could apply to different urns. This concern with heterogeneity detracts from the mixed bet on the right but is not relevant when mixing bets on the same urn as on the left. Hence the noted strict preference. (Note that \( \frac{1}{2} R_1 + \frac{1}{2} B_1 = \frac{1}{2} \), and so the indicated strict ranking is equivalent to \( W \left( \frac{1}{2} R_1 + \frac{1}{2} B_2 \right) < \frac{1}{2} \), which is readily verified.)

We consider two additional rankings that have intuitive rationales and that can be accommodated by the functional form (3.5). They are

\[
\frac{1}{2} B_1 R_2 + \frac{1}{2} R_1 B_2 \succ R_1 B_2 \sim B_1 R_2, \quad \text{and} \quad \frac{1}{2} \{B_1 B_2, R_1 R_2\} + \frac{1}{2} \{B_1 B_3, R_1 R_3\} \succ \{B_1 B_2, R_1 R_2\} \sim \{B_1 B_3, R_1 R_3\}. 
\]

The first is intuitive if the individual admits the possibility that the urns differ in composition. A good scenario for \( B_1 R_2 \) is that the first urn has more blue than red balls and the second has the opposite bias, while the opposite biases constitute a bad scenario. These “good” and “bad” scenarios are reversed for \( R_1 B_2 \). Thus \( \frac{1}{2} B_1 R_2 + \frac{1}{2} R_1 B_2 \) smooths ambiguity about differences, which motivates (5.11). The second ranking can be interpreted in terms of ambiguity about correlation. The indicated indifference is implied by Symmetry, since the two bets can be obtained from one another via the permutation that switches experiments 2 and 3.

To understand the rationale for the strict preference, suppose the individual is concerned that the urns’ compositions may follow a pattern - either the compositions of the first two are “negatively correlated”, or negative correlation exists between the first and third urns. The first pattern would make \( \{B_1 B_2, R_1 R_2\} \) a poor prospect, but not the second, and their roles reverse for \( \{B_1 B_3, R_1 R_3\} \). Therefore, randomizing between them smooths ambiguity about which pattern is valid, and leaves the mixture strictly preferable.

Verification of the latter two rankings reveals something of the differences between the two IID utility functions and our reason for including both in the specification of utility. First, \( V_{WF} \) and \( V_H \) differ in how they value the randomization
in (5.11) because
\[
V_{WF} \left( \frac{1}{2} \{B_1R_2\} + \frac{1}{2} \{R_1B_2\} \right) > V_{WF}(\{B_1R_2\}), \text{ and} \\
V_H \left( \frac{1}{2} \{B_1R_2\} + \frac{1}{2} \{R_1B_2\} \right) = V_H(\{B_1R_2\}).
\]
Secondly, with regard to the randomization in (5.12), $V_{WF}$ implies indifference to the indicated randomization, but $V_H$ values it positively. Thus both are needed in order to rationalize both (5.11) and (5.12).

The preceding rankings illustrate the scope of our model and of its specializations. The single likelihood model can accommodate (5.8) and (5.9), but it is contradicted by each of the others. The single prior model cannot accommodate (5.9). The de Finetti model cannot accommodate any of the above rankings. The present model is the only one in the literature that can accommodate all the rankings (5.8)-(5.12).

The Multiple Equilibria Example Revisited

As before take the state space $S = \{0, 1\} \times \{0, 1\}$ to describe outcomes in any single market. Any policy yields a payoff as a function of the state realized in every market $i$, where $i$ varies over $I$, which for simplicity we take to consist of the first $I$ markets, $I = \{1, 2, ..., I\}$. For example, payoffs might be constructed by aggregating the payoffs accruing from individual markets. One can then denominate payoffs in utils and normalize them to lie in the unit interval. In this way, any policy is associated with an act $f$ in $\mathcal{F}_I$, which is a subset of $\mathcal{F}$. Thus policy choice can be determined by maximizing a utility function $W$ of the form in (3.5) over the set of feasible acts. It remains to adopt and motivate a particular specification of $W$.

The presence of multiple equilibria is key. For each $\theta$ and $\epsilon$, denote the set of equilibrium mixed strategy profiles in any single market by $\Psi_\theta(\epsilon)$. Thus outcomes in any single market are summarized by the equilibrium correspondence $\Psi_\theta : \mathcal{E} \sim \Delta(S)$, for each possible $\theta$. Because the outcomes in many markets are relevant, consider the correspondence $\Psi^\infty_\theta : \mathcal{E}^\infty \sim \Delta(S^\infty)$, where
\[
\Psi^\infty_\theta(\epsilon_1, ..., \epsilon_i, ...) = \bigotimes_{i=1}^\infty \Psi_\theta(\epsilon_i) \equiv \{ \ell_1 \otimes \ell_2 \otimes ... : \ell_i \in \Psi_\theta(\epsilon_i) \text{ for all } i \}.
\]
Accordingly $\Psi^\infty_\theta(\epsilon_1, ..., \epsilon_i, ...)$ is the set of all probability distributions on the sequence of market outcomes induced by taking all possible selections of equilibria in the different markets - corresponding to the policy maker’s ignorance of the
selection mechanism and how it varies across markets - and by assuming that ran-
domizations in different markets are stochastically independent. Beliefs on $E^\infty$ are
described by the i.i.d. product measure $m^\infty$. Thus the set of possible likelihoods
for $S^\infty$ (given $\theta$) is $L_\theta^\infty \subset \Delta (S^\infty)$ given by
\[
L_\theta^\infty = \int_{E^\infty} \Psi_\theta^\infty (\varepsilon_1, \ldots, \varepsilon_i, \ldots) \, dm^\infty,
\]
the set of all mixtures, using $m$ and all possible (measurable) selections from the
sets $\Psi_\theta (\varepsilon)$.

**Remark 2.** If we define $L_\theta = \int_{E} \Psi_\theta (\varepsilon) \, dm$, then
\[
L_\theta^\infty = \int_{E^\infty} \Psi_\theta^\infty (\varepsilon_1, \ldots, \varepsilon_i, \ldots) \, dm^\infty = (L_\theta)^\infty,
\]
where the latter is defined in (5.1). The following simple case illustrates:
\[
\begin{align*}
&= \int_E \int_E \{ \ell_1 \otimes \ell_2 : \ell_i \in \Psi_\theta (\varepsilon_i) \text{ for } i = 1, 2 \} \, dm (\varepsilon_1) \, dm (\varepsilon_2) \\
&= \int_E \{ p \otimes \ell_2 : p \in L_\theta, \ell_2 \in \Psi_\theta (\varepsilon_2) \} \, dm (\varepsilon_2) \\
&= \{ p \otimes q : p, q \in L_\theta \}.
\end{align*}
\]
Thus the current specification corresponds to the IID model of beliefs in (5.1).

Finally, $\theta$ is unknown. Let beliefs about $\theta$ be given by the (possibly singleton)
set of priors $\mathcal{M} \subset \Delta (\Theta)$. Each prior $\mu$ in $\mathcal{M}$ leads to the set $\int L_\theta^\infty \, d\mu (\theta)$ of
predictive priors. Hence, for general $\mathcal{M}$, one obtains the following set $\mathcal{P}$ of predictive priors:
\[
\mathcal{P} = \bigcup_{\mu \in \mathcal{M}} \int L_\theta^\infty \, d\mu (\theta).
\]
The corresponding utility function is
\[
W (f) = \inf_{P \in \mathcal{P}} Pf = \inf_{\mu \in \mathcal{M}} \left( \inf_{P \in (L_\theta)^\infty} Pf \right) \, d\mu (\theta).
\]

\footnote{More precisely, integration of the random correspondence is in the sense of Aumann; see ES for application of Aumann integration in a similar context. If we restrict attention to the finite set $I$ of markets and if $m$ has finite support, then the corresponding integral over $E^I$ reduces to a finite summation.}
If the (mixed strategy) equilibrium is unique for all $\varepsilon$ and $\theta$ and if there is no prior ambiguity about $\theta$, then $\mathcal{P}$ reduces to a singleton consisting of a mixture of i.i.d.’s, and one obtains de Finetti’s exchangeable Bayesian model. Note that it is only through the nonsingleton sets $\mathcal{L}_q^\infty$ that our specification reflects the policy maker’s unwillingness to take a stand on the equilibrium selection mechanism and, in particular, her view that selection can vary across markets in a way that she does not understand. The exchangeable Bayesian model does not do so. Admittedly, here we are arguing based on functional forms rather than in terms of behavior, but the argument can be translated into behavioral terms as was done in the urns example.

Finally, a number of generalizations are possible. Any finite number of players is easily accommodated. So is an incomplete information game with Bayesian-Nash equilibria because multiplicity of Bayesian-Nash equilibria generates multiple likelihoods as the complete information game does. The assumption that $\varepsilon_i$ follows a particular measure $m$ can be relaxed. Instead of assuming a known distribution $m$ on $\mathcal{E}$, we could allow $m = m_\kappa$ to depend on a finite dimensional parameter $\kappa$ that would be included in $\theta$.

Another generalization is to permit differences both between markets and between players. Accordingly, suppose that payoffs in the entry game for market $i$ depend also on a variable $x_i$ that is a characteristic of the market and/or the players and is observable to both players and the analyst. For example, consider the following payoff matrix:

\[
\begin{array}{c|c|c}
\text{out} & \text{in} \\
\hline
\text{out} & 0, 0 & 0, \beta'_1 x_{1i} - \varepsilon_2 i \\
\text{in} & \beta_2 x_{2i} - \varepsilon_{1i}, 0 & \beta'_1 x_{1i} - \delta_1 - \varepsilon_{1i}, \beta_2 x_{2i} - \delta_2 - \varepsilon_{2i} \\
\end{array}
\]

The numbers $\delta_1$ and $\delta_2$ reflect the effect of competition. The variable $x_i = (x_{1i}, x_{2i})$ can represent the size of the airlines, aviation regulations or the sum of the populations of the two cities connected by market $i$. If $x_i$ includes the size of the airport, then airline payoffs are allowed to depend on airport size. For each $i$, $x_i$ lies in the finite set $X$. Now define the state space $S$ for each market by

$$S = (\{0, 1\} \times \{0, 1\})^X,$$

the set of functions from $X$ to outcomes. This specification makes intuitive sense: uncertainty regarding market $i$ concerns which outcomes will be realized for each given $x_i$. 

24
To complete the specification, assume that beliefs about \((\varepsilon_{1i}, \varepsilon_{2i})\) are described by \(m_\kappa\), where \(\kappa\) is an unknown parameter. Let \(\theta = (\beta_1, \beta_2, \delta_1, \delta_2, \kappa)\). For each \(x_i \in X\) and \(\varepsilon_i \in \mathcal{E}\), denote the set of equilibrium mixed strategy profiles by \(\Psi_{\theta, x_i}(\varepsilon_i) \subset \Delta(\{0, 1\} \times \{0, 1\})\). Thus we have the equilibrium correspondence

\[
\Psi_\theta : \mathcal{E} \rightsquigarrow (\Delta(\{0, 1\} \times \{0, 1\}))^X,
\]

where \(\Psi_\theta(\varepsilon_i) = (\Psi_{\theta, x_i}(\varepsilon_i))_{x_i}\). However, \((\Delta(\{0, 1\} \times \{0, 1\}))^X\) can be identified with a subset of \(\Delta\left(\left(\{0, 1\} \times \{0, 1\}\right)^X\right)\). (Identify \(p = (p_x)_{x \in X}, p_x \in \Delta(\{0, 1\} \times \{0, 1\}),\) with the product measure \(\otimes_{x \in X} p_x\), an element of \(\Delta\left(\left(\{0, 1\} \times \{0, 1\}\right)^X\right)\).) Thus we arrive at the equilibrium correspondence

\[
\Psi_\theta : \mathcal{E} \rightsquigarrow \Delta\left(\left(\{0, 1\} \times \{0, 1\}\right)^X\right) = \Delta(S).
\]

Finally, the set \(\mathcal{P}\) and the utility function \(W\) are constructed as above.

Just as in the simpler set up discussed above, objects of choice correspond naturally to acts. For example, a bet on the outcome \(\{(1, 0), (1, 1)\}\) in market \(i\), where \(x_i = \bar{x}\) is given, corresponds to the act \(f^i\), \(f^i : S \rightarrow [0, 1]\), where for each \(s \in S = (\{0, 1\} \times \{0, 1\})^X\),

\[
f^i(s) = \begin{cases} 
1 & \text{if } s(\bar{x}) = (1, 0) \text{ or } (1, 1) \\
0 & \text{otherwise.}
\end{cases}
\]

To illustrate the generality of our framework, consider a more complex and realistic version of the airport construction decision problem. Let \(z_{ji}\) be the size of airline \(j\) in market \(i\), let \(y_{ji}\) indicate whether airline \(j\) in market \(i\) plays in \((y_{ji} = 1)\) or out \((y_{ji} = 0)\), and denote by \(I\) the set of markets that begin or end at the new airport. Then \(H = \sum_{j \in \{1, 2\}, i \in I} y_{ji} \cdot z_{ji}\) is the total size of all airlines entering these markets. The government chooses \(d\), the size of the airport, to maximize \(u(d/H) - c(d)\) where \(u\) and \(c\) are increasing functions that represent the benefits and the costs of constructing an airport. We normalize \(u\) and \(c\) so that \(u(d/H) - c(d) \in [0, 1]\) for all possible values of \(d\) and \(H\). Another consideration is that airlines’ payoffs may depend on the size of the airport. Thus we take \(x_{ji}\) in payoff matrix (5.13) to be the vector \((z_{ji}, d)\). Now we can translate the policy of constructing an airport of size \(d\) to the act \(f_d\) from \(S^I\) to \([0, 1]\) defined by

\[
f_d(\omega) = u(d/H) - d,
\]

where \(H = \sum_{i \in I} y_{ji} \cdot z_{ji}\), and \((y_{1i}, y_{2i}) = \omega_i((z_{1i}, d), (z_{2i}, d))\).
(Again, for $\omega \in S^T$, each $\omega_i \in S$ is a function from $X$ to $\{0,1\} \times \{0,1\}$.) Given a finite feasible set of airport sizes, the choice between them is modeled by maximizing the utility function $W$ over the corresponding set of acts.

**Remark 3.** As noted in the introduction, the connection between partially identified models (such as entry games with multiple equilibria) and ambiguity averse preferences was made in Epstein and Seo (2011b). However, they exclude ambiguity about parameters and also mixed strategy equilibria.

### 6. Axiomatic Foundations

Thus far we have been discussing the functional form (3.5) for utility. Here we provide behavioral foundations for it through axioms on $\succeq$, a preference order on the set of acts $\mathcal{F}$, that lead to the representation result Theorem 6.1.

Assuming subjective expected utility, exchangeability of the predictive prior is equivalent to indifference between any act $f$ and its permuted variant $\pi f$. Thus we are led to consider, for all (not necessarily SEU) preferences, the following property that we call Symmetry:

$$f \sim \pi f \text{ for all } \pi \in \Pi \text{ and all } f \in \mathcal{F}_{fin}. $$

Symmetry in itself is a relatively weak assumption following, for example, from symmetry of information about all the experiments. We impose two axioms on preference, one of which strengthens Symmetry and thus renders it redundant.

The axiom statements are simplified by using the derived binary relation $\succeq^*$ defined on $\mathcal{F}_{fin}$ by

$$f \succeq^* g \text{ if } \lambda f + (1 - \lambda) h \succeq \lambda g + (1 - \lambda) h$$

for all $h \perp f, g$, and $\lambda \in [0, 1]$. To interpret, suppose first that the indicated ranking is satisfied for all (not necessarily orthogonal) acts $h$. Then the preference for $\lambda f + (1 - \lambda) h$ over $\lambda g + (1 - \lambda) h$ for every $h$ indicates that $f$ is weakly preferred to $g$ regardless of how either may be hedged. (This is sometimes described as “$f$ is unambiguously weakly preferred to $g$”; see Ghirardato et al. (2004) and Ghirardato and Siniscalchi (2010), for example.) In our case, invariance of the ranking is required only for acts $h$ that depend on different experiments than do $f$ and $g$. Such acts $h$ can moderate ambiguity about factors common to all experiments. Thus $f \succeq^* g$ indicates that the weak preference for $f$ over $g$ is robust to any hedging of ambiguity about parameters.
The first axiom strengthens Symmetry. If the Independence axiom is satisfied, then \( f \sim \pi f \) implies indifference also when these acts are mixed with any common third act, and this applies for any permutation \( \pi \). In order to accommodate the rankings and intuition described in the preceding section, we relax this condition by restricting it to a subset of acts and permutations (primarily by imposing suitable orthogonality).

**Axiom 1 (Weak Exchangeability (WE)).** For all finitely-based and pairwise orthogonal acts \( g, f \) and \( \pi f \), and for all \( 0 \leq \lambda \leq 1 \),

\[
\lambda f + (1 - \lambda) g \sim^* \lambda \pi f + (1 - \lambda) f \cdot g.
\]

(6.2)

Note first that WE implies Symmetry for \( \succ^* \) (a fortiori for \( \succeq^* \)): The indifference \( f \sim^* \pi f \) is a direct consequence if \( \pi f \) and \( f \) are orthogonal (take \( g = 0 \) and \( \lambda = 1 \)). More generally, given \( f \) and \( \pi \), where \( \pi f \) need not be orthogonal to \( f \), let \( \pi' \) be such that \( \pi' f \perp f, \pi f \). Then WE applied twice implies that \( f \sim^* \pi' f \sim^* \pi f \).

The axiom expresses the perception of symmetry of experiments and also that experiments are in some sense stochastically independent. The following implications of WE for preference in the urns example are telling and are helpful in justifying the preceding:\(^{20}\)

\[
R_1 \sim \alpha R_2 + (1 - \alpha) R_1,
\]

(6.3)

\[
\lambda R_2 + (1 - \lambda) B_3 \text{ and } \lambda R_1 + (1 - \lambda) B_3 \text{ do not hedge one another, and}
\]

(6.4)

\[
\alpha R_1 + (1 - \alpha) R_1 B_3 \sim \alpha R_2 + (1 - \alpha) R_1 B_3.
\]

(6.5)

One obvious implication of WE (take \( g = 1 \)) is that \( f \) and \( \pi f \) do not hedge one another if they are orthogonal, that is,

\[
\alpha f + (1 - \alpha) \pi f \sim f \text{ if } f \perp \pi f.
\]

(6.6)

Consider (6.3), for example. The intuition is firstly that \( R_1 \sim R_2 \) by Symmetry, and secondly that \( R_1 \) and \( R_2 \) pay off in similar circumstances and thus their mixture does not reduce payoff uncertainty. For each \( i = 1, 2 \), the bet \( R_i \) is likely to pay well if red is prominent in the composition of urn \( i \); and to the extent that the individual perceives a component of urn compositions that is fixed across urns, a favorable red composition occurs for both urns or for neither. In general, the

\(^{20}\)This implications of WE that follow can be derived from A.2 which gives alternative formulations that are equivalent to WE given MEU.
perception of differences between urns might support randomization. Thus an assumption implicit in (6.3), and in WE more generally, is that distinct experiments are thought to be in some sense stochastically independent. To illustrate, suppose the urns are thought to be related in the extreme sense that the individual believes that only one urn contains any red balls; the identity of the urn is ambiguous but it could equally well be any urn in the sequence. Then a red draw from one urn precludes a red draw from any other. This induces the strict preference to randomize, \( \alpha R_2 + (1 - \alpha) R_1 \succ R_1 \), because the mixed bet smooths out the uncertainty about which urn is the one containing red balls. Accordingly, indifference to randomization excludes such a perception, and more generally reveals the perception of a form of stochastic independence.

A stronger implication of WE is that, if \( h, f \) and \( \pi f \) are pairwise orthogonal, then

\[
\lambda f + (1 - \lambda) h \sim \lambda \pi f + (1 - \lambda) h
\]

and these acts do not hedge one another. The indifference in (6.4) is a special case. The two mixtures appearing there are indifferent by Symmetry. Once again, intuition for nonhedging relies on the individual being certain about a form of stochastic independence between experiments. For instance, the perception described above (it is thought that only one urn contains any red balls) motivates hedging. As an alternative example indicating violation of WE, suppose that the individual believes that all urns but two have proportion of red equal to \( \frac{1}{2} \), and also that the average proportion in the other two is \( \frac{1}{2} \). The identity of the two exceptional urns is ambiguous. The act \( \lambda R_1 + (1 - \lambda) B_3 \) pays poorly if urn 1 is the deficient urn (it has fewer than 50 red balls), and similarly for urn 2. Thus mixing them is valuable because it partially insures against the uncertainty about the identity of the deficient urn. Accordingly, the absence of hedging in (6.4) excludes such a perception.

A final implication of WE is that, if \( f, g \) and \( \pi f \) are mutually orthogonal, then

\[
\alpha f + (1 - \alpha) g \cdot f \sim \alpha(\pi f) + (1 - \alpha) g \cdot f.
\]

Aversion to ambiguity about heterogeneity (and Symmetry) would imply\(^{21}\)

\[
\alpha f + (1 - \alpha) g \cdot f \succ \alpha(\pi f) + (1 - \alpha) g \cdot f.
\]

To understand the rationale for indifference rather than strict preference, consider the special case (6.5). The intuition for indifference is that \( R_1 B_3 \) hedges \( R_1 \) and \( R_2 \)

\(^{21}\)Take \( h = g \cdot f \) and \( h' = f \) in (4.3).
equally, even though the bet $R_2$ concerns different experiments than does $R_1B_3$, while the bet $R_1$ is not orthogonal to $R_1B_3$. The reason is that the overlap in the experiments underlying $R_1$ and $R_1B_3$ is due to the common presence of $R_1$, but hedging comes only via $B_3$. Thus ambiguity about heterogeneity has no bite in this case. Similarly for (6.8).

The separate rationales for (6.7) and (6.8) can be combined to give intuition for WE.

Finally with respect to WE, we have illustrated that the axiom imposes indifference to randomization in specific cases. Instances where such indifference is not imposed by WE also reveal the connection between the axiom and the perception of experiments described in the introduction. The strict rankings (5.8)-(5.12) describe intuitive choices that are consistent with WE. In contrast, in one of their two models, ES employ an axiom called Strong Exchangeability that is simpler than WE but is too strong in that it implies the single likelihood model and thus contradicts (5.10)-(5.12).

Say that $f$ dominates $g$ if, for some $\pi_1, \pi_2 \in \Pi$, and $h, h' \in F_{\{1,\ldots,n}\}$$,$
\[ f = \pi_1 (\alpha h + (1 - \alpha) h') \quad \text{and} \quad g = \pi_2 (\alpha h + (1 - \alpha) T^n h'). \] (6.9)

It follows from (4.3) that if $f$ dominates $g$, then $f$ is weakly preferable to $g$ for any preference that is averse to ambiguity about heterogeneity and satisfies Symmetry. Denote by $f \succeq g$ the usual pointwise dominance ($f(\omega) \succeq g(\omega)$ for every $\omega$).

Our final axiom is stated next.

**Axiom 2 (Dominance).** If $f$ dominates $g$, then
\[ \frac{1}{2} f \cdot f' + \frac{1}{2} g \cdot g' \bowtie * \frac{1}{2} f \cdot g' + \frac{1}{2} g \cdot f', \] (6.10)

if also $f, g, f'$ and $g'$ are finitely-based and pairwise orthogonal and $f' \succeq \pi g'$ for some $\pi \in \Pi$.  

\[ \text{Consistency follows from the fact that } W \text{ defined by (3.5) satisfies WE (see Appendix A.1) and also accommodates the rankings (as shown in Section 5).} \]

\[ \text{Strong Exchangeability requires that } f \sim \alpha (\pi f) + (1 - \alpha) f \text{ for all } f \text{ and } \pi. \text{ The axiom’s strength is due to the indicated indifference being required for all (not necessarily orthogonal) } \pi f \text{ and } f. \text{ Closely related, and similarly excessively strong axioms appear in Najjar and de Castro (2010), Klibanoff et al. (2011) and Cerreia-Vioglio et al. (2011).} \]
If \( f \) and \( g \) are given by (6.9), and if one fixes \( f' = 1 \) and \( g' = 0 \), then (4.3) is implied (assuming Symmetry). Thus \textit{Dominance imposes a strengthening of aversion to ambiguity about heterogeneity}. One way in which the latter is strengthened is through the requirement that (4.3) be satisfied also when the preference \( \succ \) is replaced by \( \succ^* \), rendering the indicated ranking robust to any hedging of ambiguity about parameters. For further interpretation, consider (6.10) when \( \succ \) replaces \( \succ^* \).

To see another way in which aversion to ambiguity about heterogeneity is strengthened, consider the case where both \( f' = a \) and \( g' = b \), \( 0 \leq b \leq a \leq 1 \), are constant acts and where \( \alpha \) in (6.9) equals \( \frac{1}{2} \). Then, given also MEU, (6.10) implies not only that \( f \) is preferred to \( g \), but also that any mixture \( a'f + (\kappa - a')g \) increases in preference as \( a' \) increases (for any constant \( \kappa \)).

A simple example is where

\[
 f = \frac{1}{2} R_1 + \frac{1}{2} B_1 \quad \text{and} \quad g = \frac{1}{2} R_3 + \frac{1}{2} B_4,
\]

in which case \( f \) dominates \( g \) and they are orthogonal. Therefore, the axiom requires (taking \( a = \frac{2}{3} \) and \( b = \frac{1}{3} \)) that

\[
 \frac{1}{3} + \frac{1}{6} R_3 + \frac{1}{6} B_4 \succ \frac{1}{6} + \frac{1}{3} R_3 + \frac{1}{3} B_4.
\]

Doubling the certain prize from \( \frac{1}{6} \) to \( \frac{1}{3} \) more than compensates for halving the payoff from \( \frac{1}{3} R_3 + \frac{1}{3} B_4 \) to \( \frac{1}{6} R_3 + \frac{1}{6} B_4 \) because these are random (and \( \frac{1}{3} R_3 + \frac{1}{3} B_4 \) is dominated by the constant \( \frac{1}{3} \)).

Turn to the case where neither \( f' \) or \( g' \) is constant. First, assume that there is no uncertainty about factors that are common to all experiments; for example, in the entry game described in the introduction, suppose that the parameters \( \theta_1 \) and \( \theta_2 \) are known. Consistent with our story and earlier discussion, assume also that the remaining uncertainty, such as equilibrium selection in the noted entry game, is thought to be stochastically unrelated across experiments (only parameters connect experiments). Then experiments are completely unrelated, just as the outcomes of repeated tosses of a coin with known bias are i.i.d. A concrete example with ambiguity is the sequence-of-urns when the urns are perceived to have no common block of balls (\( \theta = 0 \) in the outline of Section 1.2) and to be unrelated. The behavioral meaning of the latter is expressed through rankings of the sort:

\[
 B_1 R_2 \sim p R_2 \quad \text{if} \quad B_1 \sim p.
\]

\[24\]In \( a'f + (\kappa - a')g \), the weight \( 1 - \kappa \) is attached to the constant act 0.
Because outcomes are in probability units, which permits the interpretation of product acts in Section 2, we can interpret the preceding as follows: If the bet on blue from the first urn is indifferent to betting on heads in the toss of a coin having objective probability \( p \) for heads, then the bet on blue and then red from the first two urns is indifferent to the bet that the noted objective coin lands heads and the second urn yields a red ball.

Similar intuition applies more generally and motivates the property that, for any two orthogonal acts \( f \) and \( f' \),

\[
f' \cdot f \sim a \cdot f \quad \text{if } f' \sim b.\]

Similarly, \( g' \cdot g \sim b \cdot g \) if \( g' \sim b \) and if \( g \) and \( g' \) are orthogonal. As argued in Section 4, mixing two orthogonal acts is not valuable when common factors are unambiguous (a fortiori, when they are known). Therefore,

\[
\frac{1}{2}f' \cdot f + \frac{1}{2}g \cdot g' \sim \frac{1}{2}f \cdot a + \frac{1}{2}g \cdot b,
\]

and

\[
\frac{1}{2}f \cdot g' + \frac{1}{2}g \cdot f' \sim \frac{1}{2}f \cdot b + \frac{1}{2}g \cdot a.
\]

Accordingly, (6.10) reduces to the ranking

\[
\frac{1}{2}f \cdot a + \frac{1}{2}g \cdot b \sim \frac{1}{2}f \cdot b + \frac{1}{2}g \cdot a,
\]

which was motivated above. This completes the interpretation of (6.10) in the case when all parameters (common factors) are known. But no assumptions about the values of parameters were made in the preceding. Accordingly, the ranking in (6.10) should be true more generally given any beliefs about parameters, whether probabilistic or reflecting ambiguity.

For perspective, note that (6.10) is ‘too strong’ if one requires only that \( f \succeq g \) (or \( f \approx^* g \)) in place of dominance. For instance, if \( \succeq \) is SEU, then \( \approx^* \) is and the modified axiom would require that, for all pairwise orthogonal acts \( f, g, f' \) and \( g' \), if \( f' - g' \geq 0 \), then

\[
f \succeq g \implies f \cdot (f' - g') \succeq g \cdot (f' - g').
\]

This is satisfied if the prior underlying \( \succeq \) is i.i.d. but not if it is only exchangeable.

The axiomatic foundations for our general model (3.5) can now be stated.

**Theorem 6.1.** \( \succeq \) is a MEU preference and satisfies Weak Exchangeability and Dominance if and only if it can be represented by a utility function of the form in (3.5).
7. Concluding Comments

We propose the example of a sequence of Ellsberg urns as a canonical example of ambiguity given repeated experiments. The usual way of modeling choice between bets in such a setting is to assume that the individual is certain that the urns are identical. Consequently, their common composition is the only unknown, and the exchangeable Bayesian model can be applied. But there is no reason to be confident that the urns are identical except in the isolated case where the individual is told that they are. In the entry game context the corresponding limited confidence stems from the analyst’s ignorance of the equilibrium selection mechanism, including how it might vary across markets. We expect limited confidence to matter also in many other settings where the individual realizes that her theory of the environment is incomplete, specifically because she does not understand well how experiments differ from one another.\(^{25}\) Being concerned that the outcomes of experiments may not be i.i.d., the individual seeks to make decisions that are robust to deviations from the i.i.d. specification. She also has limited prior information about the parameters that are common to all experiments and she seeks robustness also with respect to this dimension. This paper models choice that is robust in both of these senses.

Our model consists of a new functional form for utility, given by (3.5), and also its axiomatic foundations. At the functional form level, we show that its components admit natural interpretations (in part via Theorem 4.1) and that it can accommodate intuitive behavior in the sequence-of-urns example. It can be used also to model policy choice in the entry game context. At the behavioral level, we give meaning to the robustness referred to above through a series of preference rankings (recall (5.8)-(5.12)), and through the axioms Weak Exchangeability and Dominance. Besides providing intuition, these hypothetical rankings suggest new (thought and laboratory) experiments that can serve as counterparts for the setting of repeated experiments of Ellsberg’s two-color choice problem.

Given that at a formal level we are generalizing de Finetti’s theorem, it is natural to compare our representation result Theorem 6.1 with the Savage/Anscombe-Aumann/de Finetti characterization of the exchangeable Bayesian model. The main axioms in the latter are Independence and Symmetry, while in our case they are MEU, Weak Exchangeability and Dominance. We acknowledge that particularly the latter two axioms do not approach Independence and Symmetry in

\(^{25}\)Treatment choice is another context where it has been argued that such issues arise (see Manski (2011) and the references therein).
simplicity, transparency and intuitive appeal. However, the Bayesian model sets an extremely high standard that has rarely been reached or even approached in the many generalizations of expected utility pursued in the literature. We would suggest more modest criteria that are met herein. First, because the model is normative, for example, it is intended as a guide to an analyst or policy maker facing the entry game context, the decision-maker must understand the axioms well enough to decide if she finds them acceptable. Our view is that for both of the noted axioms, their essences are revealed through the illustrations provided in the sequence-of-urns context. Second, to be useful, axioms should be far removed from the representation that they imply. This is definitely the case here as evidenced by the nontrivial proof for Theorem 6.1.

This is not to say that the paper answers all relevant questions. The major limitation, in our view, is the absence of a theory of updating or inference. In contrast to our previous papers (2010, 2011b), we permit ambiguity about both parameters and heterogeneity, but at the cost of rendering inapplicable the theory of updating described there. Development of a theory of updating that can accommodate both kinds of ambiguity is the most important direction for future research.

A. Appendix: Proof of Theorem 6.1

Begin with an elementary but useful lemma.

**Lemma A.1.** Let $W$ be an MEU utility function with predictive set of priors $\mathcal{P}$. Let $f_i \in \mathcal{F}_{fin}$ and $\alpha_i > 0$, $i = 1, \ldots, n$, with $\sum_{i=1}^{n} \alpha_i = 1$. Then

$$W \left( \sum_{i=1}^{n} \alpha_i f_i \right) = \sum_{i=1}^{n} \alpha_i W \left( f_i \right)$$

if and only if there exists $P^*$ in $\mathcal{P}$ that is minimizing for every $f_i$. Further, the latter implies that, for any $m \leq n$, $\beta_i > 0$, and $\sum_{i=1}^{m} \beta_i = 1$,

$$W \left( \sum_{i=1}^{m} \beta_i f_i \right) = \sum_{i=1}^{m} \beta_i W \left( f_i \right).$$

The proof of the theorem uses extensively the derived relation $\succsim^*$ defined in (6.1). Evidently,

$$f \succsim^* g \implies f \succeq g.$$
Moreover, $\succeq^*$ is incomplete but transitive (assuming Symmetry).\footnote{Assume $f \succeq^* g$ and $g \succeq^* h$. Then, by transitivity of $\succeq$, $\alpha f + (1 - \alpha) h^* \succeq ah + (1 - \alpha) h^*$ for all $\alpha \in [0,1]$, and $h^* \perp f, g, h$. By Symmetry, the same ranking is valid for all $h^* \perp f, h$.} Also, for any $\alpha \in (0, 1)$ and $f, g, h \in F_{\text{fin}}$ such that $h \perp f, g$,

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

Say that $f^*$ hedges $f'$ and $f$ equally if $f' \sim f$ and $\alpha f' + (1 - \alpha) f^* \sim \alpha f + (1 - \alpha) f^*$ for all $\alpha$. The following equivalent restatements of WE are used below without further reference.

**Lemma A.2.** The following statements are equivalent (where $f, \pi f, g$ and $h$ vary over all finitely-based and pairwise orthogonal acts, and $0 \leq \lambda \leq 1$):

(i) $f \cdot g$ hedges $\lambda f + (1 - \lambda) h$ and $\lambda \pi f + (1 - \lambda) h$ equally.

(ii) $\lambda f \cdot g + (1 - \lambda) h$ hedges $\lambda f + (1 - \lambda) h$ and $\lambda \pi f + (1 - \lambda) h$ equally.

(iii) $\lambda f \cdot g + (1 - \lambda) h$ hedges $f$ and $\pi f$ equally.

(iv) Preference $\succeq$ satisfies WE.

**Proof.** (i) $\iff$ (ii): If

$$\alpha (\lambda f + (1 - \lambda) h) + (1 - \alpha) f \cdot g \sim \alpha (\lambda \pi f + (1 - \lambda) h) + (1 - \alpha) f \cdot g,$$

then

$$(1 - \alpha (1 - \lambda)) \left( \frac{\alpha \lambda}{1 - \alpha (1 - \lambda)} f + \frac{1 - \alpha}{1 - \alpha (1 - \lambda)} f \cdot g \right) + \alpha (1 - \lambda) h$$

$$\sim (1 - \alpha (1 - \lambda)) \left( \frac{\alpha \lambda}{1 - \alpha (1 - \lambda)} \pi f + \frac{1 - \alpha}{1 - \alpha (1 - \lambda)} f \cdot g \right) + \alpha (1 - \lambda) h.$$

Let $\lambda' = 1 - \alpha (1 - \lambda)$ and $\alpha' = \frac{\alpha \lambda}{1 - \alpha (1 - \lambda)}$, and note that $\lambda'$ and $\alpha'$ can be varied independently over $[0, 1]$. This proves $\implies$. The argument is reversible to prove the converse.

The proof that (ii) and (iii) are equivalent is similar.

(ii) $\iff$ (iv): This follows from the equalities

$$\lambda (\alpha f + (1 - \alpha) f \cdot g) + (1 - \lambda) h$$

$$= \alpha (\lambda f + (1 - \lambda) h) + (1 - \alpha) (\lambda f \cdot g + (1 - \lambda) h),$$

$$\alpha f + (1 - \alpha) h^* \succeq ah + (1 - \alpha) h^*$$

for all $\alpha \in [0,1]$, and $h^* \perp f, g, h$. By Symmetry, the same ranking is valid for all $h^* \perp f, h$.}
and
\[
\lambda (\alpha \pi f + (1 - \alpha) f \cdot g) + (1 - \lambda) h = \alpha (\lambda \pi f + (1 - \lambda) h) + (1 - \alpha) (\lambda f \cdot g + (1 - \lambda) h).
\]

Just as Symmetry was derived from WE earlier, (after the statement of WE), it is straightforward to show that (6.2) implies
\[
f \sim^* \pi f, \text{ for all } f \text{ and } \pi. \quad \text{(A.1)}
\]

**A.1. Necessity**

**MEU:** We need to show that the function $W$ in (3.5) can be expressed in the MEU form (2.1) for some convex and compact set $\mathcal{P} \subset \Delta (\Omega)$.

In ES (Appendix B.1), it is argued that: (i) $\mathcal{V}$ is compact, hence Borel measurable. (ii) $\mu$ is well-defined on the universal completion of the Borel $\sigma$-algebra on $\mathcal{V}$. (iii) The function $V \mapsto V (f)$ is universally measurable for each $f$ in $\mathcal{F}$. (iv) The integral $\int_{\mathcal{V}} V (f) \, d\mu (V)$ is well-defined for each $f$ in $\mathcal{F}$. (v) $\mathcal{Q}_{\text{IID}}$ is compact, hence Borel measurable, where
\[
\mathcal{Q}_{\text{IID}} = \{ Q \in \mathcal{K}^c (\Delta (\Omega)) : V_Q \in \mathcal{V} \},
\]
the set of all IID sets of priors, and
\[
V_Q (f) \equiv \inf_{Q \in \mathcal{Q}} Q f, f \in \mathcal{F}.
\]

(vi) Each $\mu \in \Delta (\mathcal{V})$ can be viewed as a measure on $\mathcal{K}^c (\Delta (\Omega))$, and thus we can write
\[
U (f) \equiv \int_{\mathcal{V}} V (f) \, d\mu (V) = \int V_Q (f) \, d\mu (Q). \quad \text{(A.2)}
\]

(vii) $\int Q d\mu (Q) \subset \Delta (\Omega)$ is well-defined as an Aumann integral.

Let
\[
\mathcal{P} \equiv \bigcup_{\mu \in \mathcal{M}} \int Q d\mu (Q). \quad \text{(A.3)}
\]

We show that this is the desired set of predictive priors.
We claim first that \( \mathcal{P} \) is a convex and compact subset of \( \Delta(\Omega) \). Convexity follows readily from the convexity of \( \mathcal{M} \). Compactness is proven in ES (Lemma B.3) for the special case where \( \mathcal{M} \) is a singleton. However, the proof depends only on the compactness of singletons \( \{\mu\} \) and thus applies also here for any compact set of priors \( \mathcal{M} \).

Show that \( \mathcal{P} \) is a predictive set of priors for \( W \) defined in (3.5). Use (A.2) and observe that

\[
W(f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V_{Q}(f) \, d\mu(Q) = \inf_{\mu \in \mathcal{M}} \left( \inf_{Q \in \mathcal{Q}} Qf \right) \, d\mu(Q) = \inf_{\mu \in \mathcal{M}} \inf_{P \in \mathcal{P} \mu(Q)} \int f \, dP = \inf_{P \in \mathcal{P}} \int f \, dP.
\]

**Weak Exchangeability:** First show that WE is satisfied by any IID utility function \( V \). For any two orthogonal acts \( f^* \) and \( f \),

\[
V\left(\frac{1}{4}f^* \cdot f + \frac{1}{4}f^* + \frac{1}{4}f + \frac{1}{4}\right) = V\left(\frac{1}{2}f^* + \frac{1}{2}f\right) = \frac{1}{2}V(f^*) + \frac{1}{2}V(f), \text{ if } f^* \perp f.
\]

where use has been made of the fact that \( V \) is MEU and thus satisfies “Certainty Additivity.” Apply Lemma A.1 to conclude that

\[
V\left(\frac{1}{2}f^* + \frac{1}{2}f\right) = \frac{1}{2}V(f^*) + \frac{1}{2}V(f), \text{ if } f^* \perp f.
\]

Apply this repeatedly in what follows.

Let \( \alpha, \pi f, f \) and \( h \) be pairwise orthogonal. Then

\[
V(\lambda [\alpha f + (1 - \alpha) f \cdot g] + (1 - \lambda) h) = V(\lambda f \cdot [\alpha 1 + (1 - \alpha) g] + (1 - \lambda) h)
\]

\[
= \lambda V(f \cdot [\alpha 1 + (1 - \alpha) g] + (1 - \lambda) V(h)
\]

\[
= \lambda V(f) + \lambda (1 - \alpha) V(f \cdot g) + (1 - \lambda) V(h)
\]

\[
= V(\lambda [\alpha \pi f + (1 - \alpha) f \cdot g] + (1 - \lambda) h).
\]
Now verify WE for $W$: For pairwise orthogonal acts $g, \pi f, f$ and $h$, we have

\[
W(\lambda [\alpha f + (1 - \alpha) f \cdot g] + (1 - \lambda) h) = \inf_{\mu \in M} \int_V V(\lambda [\alpha f + (1 - \alpha) f \cdot g] + (1 - \lambda) h) \, d\mu(V)
\]

\[
= \inf_{\mu \in M} \int_V V(\lambda [\alpha f + (1 - \alpha) f \cdot g] + (1 - \lambda) h) \, d\mu(V)
\]

\[
= W(\lambda [\alpha f + (1 - \alpha) f \cdot g] + (1 - \lambda) h).
\]

**Dominance:** Fix acts $f, f', g, g'$ as in the axiom. If $V$ is an IID utility function, then

\[
2V\left(\frac{1}{2} f \cdot f' + \frac{1}{2} g \cdot g'\right) - 2V\left(\frac{1}{2} f \cdot g' + \frac{1}{2} g \cdot f'\right)
= V(f \cdot f') + V(g \cdot g') - V(f \cdot g') - V(g \cdot f')
= (V(f) - V(g))(V(f') - V(g')) \geq 0.
\]

Similarly, if acts are mixed with an orthogonal act $h$. Thus, for $W$,

\[
W(\lambda [\frac{1}{2} f \cdot f' + \frac{1}{2} g \cdot g'] + (1 - \lambda) h)
= \inf_{\mu \in M} \int_V V(\lambda [\frac{1}{2} f \cdot f' + \frac{1}{2} g \cdot g'] + (1 - \lambda) h) \, d\mu(V)
\]

\[
\geq \inf_{\mu \in M} \lambda \int_V V(\lambda [\frac{1}{2} f \cdot g' + \frac{1}{2} g \cdot f'] + (1 - \lambda) h) \, d\mu(V)
\]

\[
= W(\lambda [\frac{1}{2} f \cdot g' + \frac{1}{2} g \cdot f'] + (1 - \lambda) h).
\]

**A.2. Sufficiency**

We begin with an outline of the proof of sufficiency.

By regularity, preference $\succeq$ on $\mathcal{F}$ is uniquely determined by its restriction to $\mathcal{F}_{\text{fin}}$. Because $\mathcal{F}_{\text{fin}}$ has a countable (sup-norm) dense subset $\{f_1, f_2, \ldots\}$, $\succeq$ on $\mathcal{F}_{\text{fin}}$ is uniquely determined by its restriction to $\{f_1, f_2, \ldots\}$. By Symmetry, shifting
finitely-based acts is a matter of indifference. Apply the shifting operator as many times as needed to construct \( \{T^{k_1} \tilde{f}_1, T^{k_2} \tilde{f}_2, \ldots\} \) such that any two elements are orthogonal. Now consider the restriction \( \succeq' \) of \( \succeq \) to the mixture space \( \mathcal{F}' \) generated by \( \{T^{k_1} \tilde{f}_1, T^{k_2} \tilde{f}_2, \ldots\} \). Ignoring technical details, one can mimic the argument of Gilboa and Schmeidler (1989) to show that \( \varrho' \) is represented by 

\[
W(f) = \min_{U \in C} U(f) \text{ for } f \in \mathcal{F'},
\]

for a set \( C \) of mixture linear functions on \( \mathcal{F}' \), satisfying also symmetry and (sup-norm) continuity. Moreover, any \( U \) in \( C \) can be extended to \( \mathcal{F}_{fin} \) preserving those properties. This is Lemma A.7. The preceding lemmas take care of the technical difficulties that do not arise in Gilboa and Schmeidler.

Next, we show that \( U \in C \) on \( \mathcal{F}_{fin} \) satisfies WE and Dominance as well, which is guaranteed by Lemma A.8. Suppose that \( \lambda f + (1 - \lambda) h \succeq \lambda g + (1 - \lambda) h \) for all \( \lambda \in [0, 1] \) and that \( h \perp f, g \). Then, assuming differentiability, the envelope theorem gives \( \frac{d}{d\lambda} W(\lambda f + (1 - \lambda) h) \mid_{\lambda=0} = \hat{U}(f) \) where \( \hat{U} \) solves \( \min_{U \in C} \hat{U}(h) \), and thus \( \hat{U}(f) \geq \hat{U}(g) \). Because \( h \) can vary arbitrarily, \( \hat{U} \) can be any extreme point of \( C \), and thus \( U(f) \geq U(g) \) for all \( U \in C \). Suitable specifications of \( f \) and \( g \) lead to the proof that each \( U \in C \) satisfies WE and Dominance.

Finally, it remains to show that \( U \) satisfying WE and Dominance implies that \( U \) is a single prior model. This is a reformulation of Theorem 5.2 in ES, which is described in Appendix B.

Now we prove sufficiency. Assume MEU, Weak Exchangeability and Dominance.

A preliminary observation is that some properties of preference extend to all (not necessarily finitely-based) acts in \( \mathcal{F} \). (The proof is omitted since it is analogous to similar extension results in Section 7.1 and Appendices A and B.3 of ES.)

**Lemma A.3.** (i) If \( f \succeq^* g \), where \( f, g \in \mathcal{F}_{fin} \), then \( \alpha f + (1 - \alpha) h \succeq \alpha g + (1 - \alpha) h \), for all \( h \in \mathcal{F} \) such that \( h \perp f, g \). (ii) The assertion in WE applies also to any (suitably orthogonal) act \( g \) in \( \mathcal{F} \). (iii) \( f \sim \pi f \) for every act \( f \) in \( \mathcal{F} \).

Endow \( \mathcal{F} \) with the topology induced by the sup-norm, \( \|f\| = \sup_{\omega} |f(\omega)| \); write \( f_n \to f \) if \( \|f_n - f\| \to 0 \). Since \( S \) is finite, \( \mathcal{F}_{fin} \) has the countable dense subset

\[
\{\bar{f}_1, \bar{f}_2, \ldots\} = \{f \in \mathcal{F}_{fin} : f(\omega) \in \mathbb{Q} \text{ for all } \omega \in S^\infty\}.
\]
Take a nonnegative sequence \( k_i \) of integers so that \( f_i^* \equiv T^{k_i} f_1, f_2^* \equiv T^{k_2} f_2, \ldots \) are pairwise orthogonal. Wlog, assume \( f_1^* = 0, f_2^* = 1 \) and that \( f_i^* \) is nonconstant for each \( i \geq 3 \). Define \( f_i^{**} = \frac{1}{2^i} f_i^* \) for \( i \geq 1 \). Fix \( f_1, f_2, \ldots, f_1^{**}, f_2^{**}, \ldots \) throughout the proof. Note that
\[
K \equiv \sum ||f_i^{**}|| < \infty. \tag{A.4}
\]

Let \( c_0 \) be the set of all sequences that converge to 0. Then \( c_0 \) is a Banach space under the norm \( ||\eta|| = \sup |\eta_i| \), and \( \ell_1 = \{ \eta \in \mathbb{R}^N : \sum_{i=1}^{\infty} |\eta_i| < \infty \} \) is its norm dual Aliprantis and Border (2006, Thm. 16.7). Write \( y\eta = \sum y_i \eta_i \) for \( y \in \ell_1 \) and \( \eta \in c_0 \). (We will use \( y\eta \) and \( y (\eta) \) interchangeably.) By the weak* topology on \( \ell_1 \) we mean \( \sigma (\ell_1, c_0) \), the weak topology on \( \ell_1 \) induced by \( c_0 \). Denote by \( c_0^+ \) the set of nonnegative sequences in \( c_0 \).

By the MEU assumption, \( \succeq \) on \( \mathcal{F} \) has the representation in (2.1). The latter admits an extension, also denoted \( W \), to \( \mathcal{B} (\Omega) \), the set of bounded measurable functions on \( \Omega \); \( \mathcal{B} (\Omega) \) is endowed throughout with the sup norm. The extension is simply \( W (f) = \inf_{\mathcal{P}} \int f d\mathcal{P} \) for all \( f \in \mathcal{B} (\Omega) \). Observe that
\[
W (f) = ||f|| \frac{1}{||f||} \text{ for each } f \in \mathcal{B}_+ (\Omega).
\]

Denote preference defined thereby on \( \mathcal{B} (\Omega) \) also by \( \succeq \).

Define \( \Psi : c_0 \rightarrow \mathcal{B} (\Omega) \) by
\[
\Psi (\eta) = \sum \eta_i f_i^{**}.
\]
The RHS is well-defined since \( ||\sum \eta_i f_i^{**}|| \leq ||\eta|| \sum ||f_i^{**}|| = ||\eta|| \cdot K < \infty \). Every act of the form \( \sum \eta_i f_i^{**} \) is continuous on \( \Omega \), since finite sums are continuous and converge uniformly on \( \Omega \) to the infinite sum. Note also that \( \Psi \) is continuous and one-to-one.

Define \( \succeq^\# \) on \( c_0 \) by
\[
\eta \succeq^\# \theta \text{ if } \Psi (\eta) \succeq \Psi (\theta).
\]
Then \( \succeq^\# \) is represented by the (continuous) utility function \( W^\# : c_0 \rightarrow [0, 1], \)
\[
W^\# (\eta) = W (\Psi (\eta)).
\]

Let \( B_r \) be the ball of radius \( r \) in \( \ell_1 \).
Lemma A.4. There exists $Y \subset \ell_1^+$ convex and weak* closed, such that, for any $r \geq K$ (where $K$ is defined in (A.4)), $W^\#$ can be written in the form

$$W^\#(\eta) = \min_{y \in Y \cap B_r} y\eta, \quad \eta \in c_0.$$  \hfill (A.5)

Proof. Fix $\eta \in c_0$ and hence the act $\Sigma \eta_i f_i^{**}$. Since the latter is a continuous act, and since $\mathcal{P}$ is weak-convergence compact, $W$ defined in (3.5) has a minimizing measure $P$ at any given $\Sigma \eta_i f_i^{**}$. That is, there exists $P \in \mathcal{P}$ such that

$$W(\Sigma \eta_i f_i^{**}) = P \cdot \Sigma \eta_i f_i^{**}$$

and

$$W(\Sigma \eta_i f_i^{**}) \leq P \cdot \Sigma \eta_i f_i^{**} \text{ for every } \eta' \in c_0.$$

Define $y = (y_i) = (P \cdot f_i^{**})$. Then $y$ lies in $\ell_1^+$ (since each $f_i^{**}$ is non-negative-valued and $\Sigma y_i = \Sigma P \cdot f_i^{**} \leq \Sigma \|f_i^{**}\| = K < \infty$), $W^\#(\eta) = y\eta$, and:

$$W^\#(\cdot) \leq y(\cdot) \text{ on } c_0, \text{ and}$$

$$y_2 = 1.$$  

(Recall that the second coordinate is special because it corresponds to the constant act $1$.)

Define

$$Y = \{y' \in \ell_1^+ : y'(\cdot) \geq W\#(\cdot) \text{ on } c_0, \ y'_2 = 1\}. \hfill (A.6)$$

Then, for the given $\eta$, using the fact that $\|y\| \leq K$,

$$W\#(\eta) = y\eta \geq \inf \{y'\eta : y' \in Y \cap B_K \} \geq W\#(\eta).$$

Since this is true for every $\eta$ in $c_0$, deduce that

$$W^\#(\eta) = \min_{y \in Y \cap B_K} y\eta, \text{ for all } \eta \text{ in } c_0.$$

Similarly, the equality is satisfied also if we use any ball of radius $r > K$. \hfill $\square$

Remark 4. For each $r$, $Y \cap B_r$ is convex, $\ell_1$-norm-bounded and weak* closed. Thus it is weak* compact by the Alaoglu Theorem.

Lemma A.5. For any $\eta, \eta' \in c_0^+$,
\[ [\alpha \eta + (1 - \alpha) \theta] \sim^{\#} [\alpha \eta' + (1 - \alpha) \theta] \quad \forall \alpha \in [0, 1] \quad \forall \theta \in c_0^+ \]
\[ \implies \quad [y(\eta) \geq y(\eta')] \quad \forall y \in Y \cap B_K. \]

**Proof.** \( c_0 \) is an Asplund space because its dual is separable (see Thm. 2.12 and Example 2.13 of Phelps (1989)). Moreover, as noted in the remark, \( Y \cap B_K \) is convex, and weak* compact. Thus, by Thm. 5.12 of Phelps (1989), \( Y \cap B_K \) coincides with the weak* closed convex hull of its weak* exposed points. \( \{y^\#\} \) is a weak* exposed point of \( Y \cap B_K \) if there exists \( \eta \in c_0 \) such that 
\[ \arg\min_{Y \cap B_K} y \eta = \{y^\#\}. \]

This leads us to consider an arbitrary weak* exposed point \( y^\# \) of \( Y \cap B_K \). Take \( \theta \in c_0 \) such that \( \{y^\#\} = \arg\min_{Y \cap B_K} y \theta \). Then \( \theta \in c_0^+ \): Suppose that \( \theta_k < 0 \) for some \( k \neq 2 \). Define \( y \) by \( y_i = y_i^\# \) if \( i \neq k \) and \( y_k = y_k^\# + \delta \), for \( \delta > 0 \). Then \( y \in Y \cap B_r \) for some \( r \geq K \) and \( y \theta < y^\# \theta \). But then (A.5) implies the contradiction 
\[ y^\# \theta = W^\# (\theta) \leq y \theta < y^\# \theta. \]

To exclude \( \theta_2 < 0 \), note that 
\[ y \in Y \cap B_K \implies y_2 = y_2^\# = 1 \implies y \theta - y^\# \theta \]

does not depend on the value of \( \theta_2 \). Thus we can set \( \theta_2 = 0 \) wlog.

Given that \( \theta \geq 0 \), we proceed as follows (nonnegativity is needed because the hypothesis in the lemma is restricted to \( \theta \)’s in \( c_0^+ \)). The hypothesis implies that 
\[ W^\# (\alpha \eta + (1 - \alpha) \theta) \geq W^\# (\alpha \eta' + (1 - \alpha) \theta), \]

and hence that 
\[ \frac{W^\# \left( \alpha \eta + \frac{\theta}{1 - \alpha} \right)}{W^\# \left( \frac{\alpha}{1 - \alpha} \eta + \theta \right) - W^\# (\theta)} \geq \frac{W^\# \left( \alpha \eta' + \frac{\theta}{1 - \alpha} \right)}{W^\# \left( \frac{\alpha}{1 - \alpha} \eta' + \theta \right) - W^\# (\theta)}. \]

Take \( \alpha \searrow 0 \) to derive 
\[ d^+ W^\# (\theta) (\eta) \geq d^+ W^\# (\theta) (\eta'), \]

where the directional derivative \( d^+ W^\# (\theta) (\eta) \) at \( \theta \) in direction \( \eta \) is defined by 
\[ d^+ W^\# (\theta) (\eta) = \lim_{t \searrow 0} \frac{W^\# (\theta + t \eta) - W^\# (\theta)}{t}. \]

Then, by Thm. 3 of Milgrom and Segal (2002), 
\[ \lim_{t \searrow 0} y^t (\eta) = d^+ W^\# (\theta) (\eta) \geq d^+ W^\# (\theta) (\eta') = \lim_{t \searrow 0} y'^t (\eta'), \]

41
where \( y^t \) is any element in \( \arg\min_{y \in Y} y (\theta + t \eta) \), and \( y^t \) is defined similarly.

Now show that \( \lim_{t \searrow 0} y^t (\eta) = y^\# (\eta) \). Since \((t, y) \mapsto y (\theta + t \eta)\) is continuous, \( t \mapsto \arg\min_{y \in Y} y (\theta + t \eta) \) is upper hemicontinuous by the Maximum Theorem. Moreover \( \{ y^\# \} = \arg\min_{y \in Y} y (\theta) \), and hence \( \lim_{t \searrow 0} y^t (\eta) = y^\# (\eta) \). Similarly for \( \eta' \).

Conclude that \( y^\# (\eta) \geq y^\# (\eta') \). Since \( y^\# \) is an arbitrary exposed point of \( Y \cap B_K \), \( y (\eta) \geq y (\eta') \) for all \( y \in Y \cap B_K \).

We need two special infinite permutations, \( \pi^o \) and \( \pi^e \), defined by

\[
\pi^o (i) = 2i - 1 \quad \text{for} \quad i = 1, 2, \ldots, \text{and} \\
\pi^e (i) = 2i \quad \text{for} \quad i = 1, 2, \ldots.
\]

As in Lemma A.3(iii), \( f \sim \pi^o f \) and \( f \sim \pi^e f \) for any \( f \in F \). Moreover, \( \pi^o f \perp \pi^e g \) for any \( f, g \in F \) since \( \pi^o f \) depends only on odd numbered experiments and \( \pi^e g \) only on even numbered experiments.

**Lemma A.6.** For all \( \alpha \in [0, 1] \) and \( \eta, \theta \in c_0^+ \),

\[
\alpha \Psi (\eta) + (1 - \alpha) \Psi (\theta) \sim \alpha \pi^o \Psi (\eta) + (1 - \alpha) \pi^e \Psi (\theta).
\]

**Proof.** By WE and Lemma A.3,

\[
\alpha \Psi (\eta) + (1 - \alpha) \Psi (\theta) \sim \pi^e (\alpha \Psi (\eta) + (1 - \alpha) \Psi (\theta)) \\
= \alpha \pi^e \left( \sum_{i=1}^{\infty} \eta_i f_i^{**} + (1 - \alpha) \pi^e \sum_{i=1}^{\infty} \theta_i f_i^{**} \right) \\
\sim \alpha \left[ \pi^o \sum_{i=1}^{3} \eta_i f_i^{**} + \pi^e \sum_{i=4}^{\infty} \eta_i f_i^{**} \right] + (1 - \alpha) \pi^e \sum_{i=1}^{\infty} \theta_i f_i^{**} \quad (I) \\
\sim \ldots \sim \alpha \left[ \pi^o \sum_{i=1}^{n} \eta_i f_i^{**} + \pi^e \sum_{i=n+1}^{\infty} \eta_i f_i^{**} \right] + (1 - \alpha) \pi^e \sum_{i=1}^{\infty} \theta_i f_i^{**}.
\]

We elaborate on the proof of the indifference (I) - the argument for subsequent indifferences is similar.

Let \( g = 1 \) in WE. Then, by Lemma A.3,

\[
\lambda f + (1 - \lambda) h \sim \lambda [\alpha \pi f + (1 - \alpha) f] + (1 - \lambda) h,
\]
whenever $f, \pi f, h$ are pairwise orthogonal, $f \in F_{fin}$ and $h \in F$ is allowed to be non-finitely-based. This is applied as follows to prove (I). Note that

$$
\alpha \pi \sum_{i=1}^{\infty} \eta_i f^{**}_i + (1 - \alpha) \pi \sum_{i=1}^{\infty} \theta_i f^{**}_i
= (\alpha \eta_3 + (1 - \alpha) \theta_3) (\pi e f_3^{**}) + \alpha \pi \sum_{i=1, i \neq 3}^{\infty} \eta_i f^{**}_i + (1 - \alpha) \pi \sum_{i=1, i \neq 3}^{\infty} \theta_i f^{**}_i.
$$

Both $\pi^o f_3^{**}$ and $\pi^e f_3^{**}$ are orthogonal to $\alpha \pi \sum_{i=1, i \neq 3}^{\infty} \eta_i f^{**}_i + (1 - \alpha) \pi \sum_{i=1, i \neq 3}^{\infty} \theta_i f^{**}_i$. Thus,

$$
(\alpha \eta_3 + (1 - \alpha) \theta_3) (\pi e f_3^{**}) + \alpha \pi \sum_{i=1, i \neq 3}^{\infty} \eta_i f^{**}_i + (1 - \alpha) \pi \sum_{i=1, i \neq 3}^{\infty} \theta_i f^{**}_i
\sim \alpha \eta_3 (\pi^o f_3^{**}) + (1 - \alpha) \theta_3 (\pi^e f_3^{**}) + \alpha \pi \sum_{i=1, i \neq 3}^{\infty} \eta_i f^{**}_i + (1 - \alpha) \pi \sum_{i=1, i \neq 3}^{\infty} \theta_i f^{**}_i
= \alpha \left[ \pi^o \eta_3 f_3^{**} + \pi^e \sum_{i=1, i \neq 3}^{\infty} \eta_i f^{**}_i \right] + (1 - \alpha) \pi \sum_{i=1}^{\infty} \theta_i f^{**}_i.
$$

Since $f_1^{**}$ and $f_2^{**}$ are constant, (I) obtains.

Finally, observe that

$$
\left\| \left( \pi^o \sum_{i=1}^{n} \eta_i f_i^{**} + \pi^e \sum_{i=n+1}^{\infty} \eta_i f_i^{**} \right) - \left( \pi^o \sum_{i=1}^{\infty} \eta_i f_i^{**} \right) \right\| \leq \sum_{i=n+1}^{\infty} \eta_i \| \pi^e f_i^{**} - \pi^o f_i^{**} \| \to 0
$$

as $n \to \infty$, because $\sum_{i=1}^{\infty} \eta_i \| \pi^e f_i^{**} - \pi^o f_i^{**} \| < \infty$. Thus norm-continuity of $W$ completes the proof. \hfill \square

Let $\Pi^* = \Pi \cup \{T^k : k = 1, 2, \ldots\}$, the set of all finite permutations and finite shift operations. Define

$$
\mathcal{G} \equiv \bigcup \{ \pi \Psi \left( c_0^\gamma \right) : \pi^* \in \Pi^* \}, \text{ and}
\mathcal{H} \equiv \{ \alpha f + (1 - \alpha) g : f \in \mathcal{G}, g \in F_{fin}, 0 \leq \alpha \leq 1 \}.
$$

**Lemma A.7.** There is a set $C$ of functions $U$ defined on $\mathcal{H}$, such that

$$
W(f) = \min_{U \in C} U(f) \text{ on } \mathcal{H},
$$

43
and for all $U \in C$: i) $U (\alpha f + (1 - \alpha) g) = \alpha U (f) + (1 - \alpha) U (g)$ for all orthogonal $f \in \mathcal{G}$ and $g \in \mathcal{F}_{fin}$; ii) $U (f) = U (\pi f)$ for all $\pi \in \Pi$ and $f \in \mathcal{F}_{fin}$; iii) $U (c) = c$ for all $c \in [0, 1]$; iv) $U (f) - U (g) \leq \|f - g\|$ for all $f, g \in \mathcal{H}$.

**Proof.** For each $y \in Y \cap B_K$, define $U_y$ on $\Psi (c_0^+) = \{ \Psi (\eta) : \eta \in c_0^+ \}$ by

$$U_y (f) = y \Psi^{-1} (f).$$

Let

$$C = \{ U_y : y \in Y \cap B_K \}.$$  \hspace{1cm} (A.7)

It follows from (A.5) that $\mathcal{Y}$ can be represented by $f \mapsto \min_{U \in C} U (f)$ when restricted to $\Psi (c_0^+)$, where it must therefore be ordinally equivalent to $W$. But $U (c) = c$ for all $c \geq 0$, and $W$ is normalized in the same way. Therefore,

$$W (f) = \min_{U \in C} U (f) \text{ for } f \in \Psi (c_0^+).$$

The main task is to extend this equality suitably to $\mathcal{H}$.

We prove the following which will be useful below: for any $f, g \in \mathcal{F}$, $\pi^* \in \Pi^*$ and $0 \leq \alpha \leq 1$,

$$\alpha \pi^o f + (1 - \alpha) \pi^e g \sim \alpha \pi^o \pi^* f + (1 - \alpha) \pi^e g.$$  \hspace{1cm} (A.8)

Since both $\pi^o f$ and $\pi^o \pi^* f$ are orthogonal to $\pi^e g$, there exists a (not necessarily finite) permutation $\pi$ such that

$$\pi [\alpha \pi^o f + (1 - \alpha) \pi^e g] = \alpha \pi^o \pi^* f + (1 - \alpha) \pi^e g.$$  

If $f$ is finite, $\pi$ can be taken to be a finite permutation, and then (A.8) holds. Thus, the two functions, $f \mapsto W (\alpha \pi^o f + (1 - \alpha) \pi^e g)$ and $f \mapsto W (\alpha \pi^o \pi^* f + (1 - \alpha) \pi^e g)$, coincide on $\mathcal{F}_{fin}$. By the extension arguments mentioned prior to Lemma A.3, equality extends to all of $\mathcal{F}$.

Next we show that, for all $U$ in $C$,

$$U (\pi^* f) = U (f) \text{ if } f, \pi^* f \in \Psi (c_0^+).$$

Let $f = \Psi (\eta)$ and $\pi^* f = \Psi (\eta')$. Then, by Lemmas A.6 and A.3,

$$\alpha \Psi (\eta) + (1 - \alpha) \Psi (\theta) \sim \alpha \pi^o \Psi (\eta) + (1 - \alpha) \pi^o \Psi (\theta)$$

$$\sim \alpha \pi^o \pi^* \Psi (\eta) + (1 - \alpha) \pi^e \Psi (\theta) \quad \text{(by (A.8))}$$

$$= \alpha \pi^o \Psi (\eta') + (1 - \alpha) \pi^e \Psi (\theta)$$

$$\sim \alpha \Psi (\eta') + (1 - \alpha) \Psi (\theta).$$

44
or \( \alpha y + (1 - \alpha) \theta \sim \# \alpha y' + (1 - \alpha) \theta \). Since this is true for all \( \alpha \) and \( \theta \in c_0^+ \),

\( y \eta = y \eta' \) for all \( y \in Y \cap B_K \) by Lemma A.5. Hence \( U(f) = U(\pi f) \) for all \( U \in C \).

Extend \( U \) from \( \Psi(c_0^+) \) to \( G \) by

\[
U(\pi^* f) = U(f) \quad \text{for } f \in \Psi(c_0^+) \quad \text{and } \pi^* \in \Pi^*.
\]

By the preceding, \( U \) is well-defined.

Next extend \( U \) to \( H \). To do so, recall that \( \{\tilde{f}_1, \tilde{f}_2, \ldots\} \) is a countable dense subset of \( \mathcal{F}_{\text{fin}} \), and note that \( \{\tilde{f}_1, \tilde{f}_2, \ldots\} \subset G \). Moreover, for any \( f \in H \), there is a sequence \( f_n \in \mathcal{F}_{\text{fin}} \) that converges to \( f \); if \( f = \alpha \pi^* \Psi(\eta) + (1 - \alpha) g \in H \) with \( \eta \in c_0^+, \pi^* \in \Pi^* \) and \( g \in \mathcal{F}_{\text{fin}} \), then we can approximate \( \Psi(\eta) \) by \( \Psi(\eta^n) \) where \( \eta^n \in c_0^+ \) is defined by \( \eta^n_i = \eta_i \) if \( i \leq n \) and 0 otherwise. Therefore, the desired (norm-continuous) extension to \( H \) exists and it is unique if

\[
U(f) - U(g) \leq \|f - g\| \quad \text{for every } f, g \in G.
\]

Prove the latter in steps.

Step 1. Assume \( f \geq f', f, f' \in G \), and show that \( U(f) \geq U(f') \) for any \( U \in C \):

Note that \( \pi f = \Psi(\eta) \) and \( \pi' f' = \Psi(\eta') \) for some \( \pi, \pi' \in \Pi^* \), \( \eta, \eta' \in c_0^+ \). Take arbitrary \( \alpha \in [0, 1] \) and \( \theta \in c_0^+ \). Then Lemmas A.6 and A.3 imply:

\[
\alpha \Psi(\eta) + (1 - \alpha) \Psi(\theta) \sim \alpha \pi^0 \Psi(\eta) + (1 - \alpha) \pi^e \Psi(\theta) \\
= \alpha \pi^0 \pi f + (1 - \alpha) \pi^e \Psi(\theta) \\
\sim \alpha \pi^0 \pi' f + (1 - \alpha) \pi^e \Psi(\theta) \quad \text{(by A.8)} \\
\geq \alpha \pi^0 \Psi(\eta') + (1 - \alpha) \pi^e \Psi(\theta) \sim \alpha \Psi(\eta') + (1 - \alpha) \Psi(\theta);
\]

in other words, \( \alpha \eta + (1 - \alpha) \theta \sim \# \alpha \eta' + (1 - \alpha) \theta \), and this is true for each \( \theta \) in \( c_0^+ \).

By Lemma A.5, \( y(\eta) \geq y(\eta') \) for each \( y \in Y \cap B_K \). Conclude, by the definition (A.7) of \( C \), that, for each \( U \) in \( C \),

\[
U(f) = U(\pi f) = y(\eta) \geq y(\eta') = U(\pi' f') = U(f').
\]

Step 2. Take any \( f, g \) in \( G \), and note that \( f \leq g + \|f - g\| \). Clearly, \( g + \|f - g\| \)

lies in \( G \) (\( \|f - g\| \) is a constant and the second coordinate of \( \eta \) represents 1). By

Step 1, the desired inequality (A.9) is satisfied.
We now have each $U$ in $C$ well-defined (and norm-continuous) on $\mathcal{H}$. Next show that every such function $U$ satisfies the stated properties.

i) $U(\alpha f + (1 - \alpha) g) = \alpha U(f) + (1 - \alpha) U(g)$ for all orthogonal $f \in \mathcal{G}$ and $g \in \mathcal{F}_{\text{fin}}$: Take $y \in Y \cap B_K$ such that $U = U_y$ on $\Psi(c^+_0)$.

First, assume $f, g \in \mathcal{F}_{\text{fin}}$ and wlog let $f \in \mathcal{F}_{\{1,...,k\}}$ and $g \in \mathcal{F}_{\{k+1,...,2k\}}$. Recall that $\{\tilde{f}_1, \tilde{f}_2, \ldots\}$ is the set of all acts $f$ in $\mathcal{F}_{\text{fin}}$ such that $f(\omega)$ is rational for each $\omega \in \Omega$. Thus, we can take sequences $f_n \in \{\tilde{f}_1, \tilde{f}_2, \ldots\} \cap \mathcal{F}_{\{1,...,k\}}$ and $g_n \in \{\tilde{f}_1, \tilde{f}_2, \ldots\} \cap \mathcal{F}_{\{k+1,...,2k\}}$ that converge to $f$ and $g$ respectively. In particular, $f_n \perp g_n$. For each $n$, take $\eta^n, \theta^n \in c^+_0$ and $\pi_n, \pi'_n \in \Pi$ such that $f_n = \pi_n \Psi(\eta^n)$ and $g_n = \pi'_n \Psi(\theta^n)$. Note that, for each $n$, $\eta^n_i > 0$ and $\theta^n_j > 0$ for only one $i$ and $j$, and $i \neq j$. Thus, $\Psi(\eta^n) \perp \Psi(\theta^n)$ as well and thus there exists $\pi''_n \in \Pi$ such that $\alpha f_n + (1 - \alpha) g_n = \pi''_n (\alpha \Psi(\eta^n) + (1 - \alpha) \Psi(\theta^n))$. Observe that

$$U(\alpha f_n + (1 - \alpha) g_n) = U(\alpha \Psi(\eta^n) + (1 - \alpha) \Psi(\theta^n)) = y(\alpha \eta^n + (1 - \alpha) \theta^n) = \alpha y(\eta^n) + (1 - \alpha) y(\theta^n) = \alpha U(f_n) + (1 - \alpha) U(g_n).$$

Taking the limit gives the desired result.

Now, let $f \in \mathcal{G}$ and $g \in \mathcal{F}_{\text{fin}}$. Then $f = \pi^* \Psi(\eta)$ for some $\pi^* \in \Pi^*$ and $\eta \in c^+_0$. For each $n$, define $\eta^n \in c^+_0$ by $\eta^n_i = \eta_i$ if $i \leq n$, 0 otherwise. Then, $f_n = \Psi(\eta^n)$ is a sequence in $\mathcal{F}_{\text{fin}}$ that converges to $f$. By the preceding case,

$$U(\alpha f_n + (1 - \alpha) g) = \alpha U(f_n) + (1 - \alpha) U(g)$$

for each $n$. Take the limit to get the desired result.

ii) $U(f) = U(\pi f)$ for all $\pi \in \Pi$ and $f \in \mathcal{F}_{\text{fin}}$: $U(f) = U(\pi f)$ for all $f \in \mathcal{G}$, and the norm closure of $\mathcal{G}$ contains $\mathcal{F}_{\text{fin}}$.

iii) $U(c) = c$ for all $c \in [0, 1]$: Clear.

iv) $U(f) - U(g) \leq ||f - g||$ for all $f, g \in \mathcal{H}$: Clear.

Next, show that $\min_{U \in \mathcal{C}} U(f)$ is well-defined for $f \in \mathcal{H}$. Consider 3 cases.

Case 1. $f \in \mathcal{G}$: Then $\pi^* f = \Psi(\eta)$ for some $\pi^* \in \Pi$ and $\eta \in c^+_0$. Since $Y \cap B_K$ is weak* compact, $\min_{y \in Y \cap B_K} y \eta$ exists and it equals $\min_{U \in \mathcal{C}} U(f)$.

Case 2. $f \in \mathcal{F}_{\text{fin}}$: Let $f_n \to f$ where $f_n \in \mathcal{G}$. Then, there is a minimizer $U_n$ for each $f_n$, i.e., $U_n \in \text{argmin}_{U \in \mathcal{C}} U(f_n)$. Take $y_n \in Y \cap B_K$ such that $U_n(\cdot) = y_n \Psi^{-1}(\cdot)$ on $\Psi(c^+_0)$. Since $Y \cap B_K$ is weak* compact, wlog assume $y_n \to \hat{y} \in Y \cap B_K$.
\( B_K \). Let \( \hat{U} (\cdot) = \hat{y} \Psi^{-1} (\cdot) \) on \( \Psi \left( c_0^+ \right) \), and recall that the same notation \( \hat{U} \) denotes the extension to \( \mathcal{H} \). Show that \( \hat{U} \) is a minimizer for \( f \), that is, \( \hat{U} (f) \leq U (f) \) for all \( U \in C \). Suppose that \( \hat{U} (f) > U^* (f) \) for some \( U^* \in C \). Let \( \varepsilon = \hat{U} (f) - U^* (f) > 0 \).

Since \( \| f_n - f \| \to 0 \), take an integer \( N \) such that \( \| f_n - f \| < \frac{\varepsilon}{5} \) for all \( n \geq N \).

Since \( f_n \in \mathcal{G} \), \( f_N = \pi^* \Psi (\eta) \) for some \( \pi^* \in \Pi^* \) and \( \eta \in c_0^+ \). There exists an integer \( M \) such that \( \| U_n (f_N) - \hat{U} (f_N) \| = \| y_n (\eta) - \hat{y} (\eta) \| < \frac{\varepsilon}{5} \) for all \( n \geq M \). Set \( n = \max \{ N, M \} \) and observe that

\[
\begin{align*}
\left| U_n (f_n) - \hat{U} (f) \right| & \leq \left| U_n (f_n) - U_n (f) \right| + \left| U_n (f) - U_n (f_N) \right| + \left| U_n (f_N) - \hat{U} (f_N) \right| \\
& \quad + \left| \hat{U} (f_N) - \hat{U} (f) \right| \\
& \leq \| f_n - f \| + 2 \| f_N - f \| + \left| U_n (f_N) - \hat{U} (f_N) \right| < \frac{4}{5} \varepsilon,
\end{align*}
\]

where the second inequality holds because of iv). Moreover,

\[
\left| U^* (f_n) - U^* (f) \right| \leq \| f_n - f \| < \frac{1}{5} \varepsilon.
\]

Conclude that

\[
U_n (f_n) > \hat{U} (f) - \frac{4}{5} \varepsilon = U^* (f) + \frac{1}{5} \varepsilon > U^* (f_n),
\]

which contradicts that \( U_n \) is a minimizer for \( f_n \).

Before proceeding to Case 3, we show that

\[
\text{if } y_n \to y, \text{ then } U_{y_n} (f) \to U_y (f) \text{ for each } f \in \mathcal{F}_{fin}. \tag{A.10}
\]

Here is a proof: Take any \( \varepsilon > 0 \) and choose \( f' = \pi^* \Psi (\eta) \) for some \( \eta \in c_0^+ \), such that \( \| f - f' \| < \varepsilon \). There exists \( N_0 \) such that, for all \( n \geq N_0 \), \( \| U_{y_n} (f') - U_y (f') \| = \| y_n (\eta) - y (\eta) \| < \varepsilon \), and hence

\[
\begin{align*}
\left| U_{y_n} (f) - U_y (f) \right| & \leq \left| U_{y_n} (f) - U_{y_n} (f') \right| + \left| U_{y_n} (f') - U_y (f') \right| + \left| U_y (f') - U_y (f) \right| \\
& \leq 2 \| f - f' \| + \left| U_{y_n} (f') - U_y (f') \right| < 3 \varepsilon.
\end{align*}
\]

Case 3. \( f \in \mathcal{H} \): Take \( f_n \in \mathcal{F}_{fin} \) that converges to \( f \), and argue as in Case 2. Each \( f_n \) has a minimizer \( U_n \), a corresponding element \( y_n \in Y \cap B_K \); and wlog \( y_n \to \hat{y} \in Y \cap B_K \); \( \hat{y} \) defines \( \hat{U} \in C \). Show that \( \hat{U} \) is a minimizer for \( f \). Suppose that \( \hat{U} (f) > U^* (f) \) for some \( U^* \in C \). Let \( \varepsilon = \hat{U} (f) - U^* (f) > 0 \). Since
\( \|f_n - f\| \to 0 \), take an integer \( N \) such that \( \|f_n - f\| < \frac{\varepsilon}{5} \) for all \( n \geq N \). By (A.10), we can take an integer \( M \) such that \( \left| U_n (f_N) - \hat{U} (f_N) \right| < \frac{\varepsilon}{5} \) for all \( n \geq M \). Set \( n = \max \{ N, M \} \) and observe that

\[
\begin{align*}
\left| U_n (f_n) - \hat{U} (f) \right| & \leq |U_n (f_n) - U_n (f)| + |U_n (f) - U_n (f_N)| + |U_n (f_N) - \hat{U} (f_N)| \\
& \quad + \left| \hat{U} (f_N) - \hat{U} (f) \right| \\
& \leq \|f_n - f\| + 2 \|f_N - f\| + \left| U_n (f_N) - \hat{U} (f_N) \right| < \frac{1}{5} \varepsilon,
\end{align*}
\]

where the second inequality holds because of iv). Moreover,

\[
|U^* (f_n) - U^* (f)| \leq \|f_n - f\| < \frac{1}{5} \varepsilon.
\]

Conclude that

\[
U_n (f_n) > \hat{U} (f) - \frac{1}{5} \varepsilon = U^* (f) + \frac{1}{5} \varepsilon > U^* (f_n),
\]

which contradicts that \( U_n \) is a minimizer for \( f_n \).

Finally, show that \( W (\cdot) = \min_{U \in C} U (\cdot) \) on \( \mathcal{H} \). The two functions coincide on \( \Psi (c_0^+) \) and, hence, by Lemma A.3, also on \( \mathcal{G} \). Finally, both are norm-continuous and the norm-continuous extension from \( \mathcal{G} \) to \( \mathcal{H} \) is unique. \( \square \)

**Lemma A.8.** If \( f \gtrsim g \) for \( f, g \in \mathcal{F}_{fin} \), then \( U (f) \geq U (g) \) for all \( U \in C \).

**Proof.** The proof is essentially the same as that of Lemma A.5.

Take any weak* exposed point \( y^* \) of \( Y \cap B_K \), and let \( h \in \Psi (c_0) \) be such that \( \{ U_{y^*} \} = \text{argmin}_{U \in C} U (h) \). Argue as in the proof of Lemma A.5 that \( h \in \Psi (c_0^+) \). By shifting \( h \) suitably, it is wlog to assume that \( h \perp f, g \) and \( h \in \mathcal{G} \). Therefore, \( f \gtrsim^* g \) implies

\[
W (\lambda f + (1 - \lambda) h) \geq W (\lambda g + (1 - \lambda) h),
\]

for all \( \lambda \in (0, 1) \). Hence,

\[
W \left( \frac{\lambda}{1 - \lambda} f + h \right) \geq W \left( \frac{\lambda}{1 - \lambda} g + h \right) \quad \text{and} \quad \frac{\frac{\lambda}{1 - \lambda} f + h - W (h)}{\frac{\lambda}{1 - \lambda}} \geq \frac{\frac{\lambda}{1 - \lambda} g + h - W (h)}{\frac{\lambda}{1 - \lambda}}.
\]
Take $\lambda \not\subset 0$ to derive $d^+W (h) (f) \geq d^+W (h) (g)$, where the directional derivative $d^+W$ is defined by

$$d^+W (h)(f) = \lim_{t\searrow 0} \frac{W (h + tf) - W (h)}{t}.$$ 

Then, by Thm. 3 of Milgrom and Segal (2002),

$$\lim_{t\searrow 0} U^t_f (f) = d^+W (h)(f) \geq d^+W (h)(g) = \lim_{t\searrow 0} U^t_g (g),$$

where $U^t_f$ is any element in $\text{argmin}_{U \in Y} U (h + tf)$, and $U^t_g$ is defined similarly. We know that $U (h + tf) = U (h) + tU (f)$ by Lemma A.7.

Identify $C$ with $Y \cap B_K \subset \ell_1$ as in (A.7), where $U = U_y$ is identified with $y$. Equip $C$ with the topology induced via this identification by the weak* topology on $Y \cap B_K$. This topology is metrizable ($c_0$ is separable and we can apply Thm. 6.30 of Aliprantis and Border (2006).)

Claim: For any sequence $U_n$ in $C$ converging to $U$, $U_n (f) \to U (f)$ for every $f$ in $H$. When $f \in F_{\text{fin}}$, the claim is proven in (A.10). Essentially the same proof works for $f \in H$. To see that, let $f \in H$. Take $\varepsilon > 0$ and $f' \in F_{\text{fin}}$ such that $\|f - f'\| < \varepsilon$. By (A.10), there is $N_0$ such that $|U_n (f') - U (f')| < \varepsilon$ for all $n \geq N_0$. Then, for all $n \geq N_0$,

$$|U_n (f) - U (f)| \leq |U_n (f) - U_n (f')| + |U_n (f') - U (f')| + |U (f') - U (f)| \leq 2\|f - f'\| + |U_n (f') - U (f')| < 3\varepsilon.$$ 

Now show that $\lim_{t\searrow 0} U^t_f (f) = U^*_y (f)$. By the Claim, it is straightforward to show that $(t, U) \longmapsto U (h) + tU (f)$ is continuous. By the Maximum Theorem, $t \longmapsto \text{argmin}_{U \in C} U (h) + tU (f)$ is upper hemicontinuous. Moreover $\{U^*_y\} = \text{argmin}_{U \in C} U (h)$, and hence $U^*_f \to U^*_y$, as $t \searrow 0$. By the Claim, $\lim_{t\searrow 0} U^t_f (f) = U^*_y (f)$. Similarily, $\lim_{t\searrow 0} U^t_g (g) = U^*_y (g)$.

Conclude that $U^*_y (f) \geq U^*_y (g)$. Since $y^*$ is an arbitrary weak* exposed point of $Y \cap B_K$, $U (f) \geq U (g)$ for all exposed points $U$ of $C$. By Thm. 5.12 of Phelps (1989), $U (f) \geq U (g)$ for all $U \in C$. \hfill \Box

By the previous lemma, $U \in C$ inherits Dominance. In addition, by Lemma A.7, WE implies

$$U (\alpha f + (1 - \alpha) f \cdot g) = U (\alpha pf + (1 - \alpha) f \cdot g) = \alpha U (pf) + (1 - \alpha) U (f \cdot g) = \alpha U (f) + (1 - \alpha) U (f \cdot g),$$

for each $U \in C$ and orthogonal $f, g \in F_{\text{fin}}$. 

49
Lemma A.9. For each $U \in C$, $U (f \cdot (\alpha 1 + (1 - \alpha) g)) = \alpha U (f) + (1 - \alpha) U (f \cdot g)$ for all orthogonal $f, g \in \mathcal{F}_{fin}$, and $U$ satisfies Symmetry and Dominance.

Return to the proof of the theorem. Each $U$ in $C$ is defined on $\mathcal{F}_{fin}$. By Thm. D.2 of Epstein and Wang (1995), there is a unique regular extension to $\mathcal{F}$, which is still denoted by $U$. We have shown that $W (f) = \inf_{U \in C} U (f)$ represents $\succsim$ on $\mathcal{F}_{fin}$. In addition, the preceding lemma shows that each $U$ in $C$ satisfies the axioms characterizing the single prior model as described in the next appendix (Theorem B.1).\(^{27}\) Thus, for each $U$ in $C$, there exists $\mu_U \in \Delta (\mathcal{V})$ such that

$$
U (f) = \int_{\mathcal{V}} V (f) d\mu_U (V) \text{ for every } f \in \mathcal{F}.
$$

The set $\{\mu_U : U \in C\}$ is convex (the mapping $U \rightarrow \mu_U$ is mixture linear). Let $\mathcal{M}$ be its weak-convergence closure, which is also convex, and define

$$
\tilde{W} (f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V (f) d\mu (V) \text{ for each } f \in \mathcal{F}.
$$

Show that $W = \tilde{W}$ on $\mathcal{F}_{fin}$. Let $f \in \mathcal{F}_{fin}$. Since $V \mapsto V (f)$ is continuous (by the Maximum Theorem), $\mu \mapsto \int_{\mathcal{V}} V (f) d\mu (V)$ is continuous and hence,

$$
\tilde{W} (f) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V (f) d\mu (V) = \inf_{\mu \in \{\mu_U : U \in C\}} \int_{\mathcal{V}} V (f) d\mu_U (V) = \inf_{U \in C} U (f) = W (f), \text{ for every } f \in \mathcal{F}_{fin}.
$$

Then, by Lemma 2.1, $W = \tilde{W}$ on $\mathcal{F}$, which completes the proof.

**B. The Single Prior Model**

We provide a representation theorem for the single prior model that is slightly different from Thm. 5.2 in ES and that is more convenient for the proof of Theorem 6.1.

\(^{27}\)The theorem is a reformulation of Thm. 5.2 in ES.
Assume that the MEU function $U$ represents $\succsim$. Say that $\succsim$ satisfies Orthogonal Independence (OI) if

$$U(\alpha f + (1 - \alpha) f \cdot g) = \alpha U(f) + (1 - \alpha) U(f \cdot g),$$

for all orthogonal $f, g \in \mathcal{F}_{fin}$ and $0 \leq \alpha \leq 1$. In light of (2.2), it is not difficult to express OI in terms of behavior. One implication of OI is:

$$U \left( \sum_{\emptyset \neq I \subset \{1, \ldots, n\}} \alpha_I \left( \prod_{i \in I} f_i \right) \right) = \sum_{\emptyset \neq I \subset \{1, \ldots, n\}} \alpha_I U \left( \prod_{i \in I} f_i \right) \quad (B.1)$$

if $f_i$’s are pairwise orthogonal, $\alpha_I \geq 0$ and $\sum_{\emptyset \neq I \subset \{1, \ldots, n\}} \alpha_I = 1$. To see this, apply OI to compute that

$$U \left( \left( \frac{1}{2} + \frac{1}{2} f_1 \right) \left( \frac{1}{2} + \frac{1}{2} f_2 \right) \right) = \frac{1}{2} U \left( \frac{1}{2} + \frac{1}{2} f_1 \right) + \frac{1}{2} U \left( \left( \frac{1}{2} + \frac{1}{2} f_1 \right) \cdot f_2 \right) = \frac{1}{2} \left[ \frac{1}{2} + \frac{1}{2} U(f_1) \right] + \frac{1}{2} \left[ \frac{1}{2} U(f_2) + \frac{1}{2} U(f_1 \cdot f_2) \right]$$

and thus $f_1, f_2$ and $f_1 \cdot f_2$ have a common minimizing measure. Argue similarly for any product of the form $\left( \frac{1}{2} + \frac{1}{2} f_1 \right) \cdots \left( \frac{1}{2} + \frac{1}{2} f_n \right)$ and conclude that there is a common minimizing measure for all terms obtained in its expansion. Since $U$ is MEU, (B.1) is implied by Lemma A.1. A special case of (B.1) is

$$U \left( \sum \alpha_i f_i \right) = \sum \alpha_i U(f_i),$$

for pairwise orthogonal $f_1, \ldots, f_n$.

**Theorem B.1.** $\succsim$ is a MEU preference satisfying Symmetry, OI and Dominance if and only if it can be represented by

$$U(f) = \int_V V(f) d\mu(V), \ f \in \mathcal{F},$$

for some (necessarily unique) $\mu \in \Delta(V)$.

**Proof.** For necessity, Symmetry and OI are readily verified. Both Dominance and MEU are verified in Appendix A.1. Uniqueness is proven in our previous paper, and the proof of sufficiency for Theorem 5.2 of the latter paper can be adapted to the present reformulation. We sketch the new proof focusing on the differences from the previous argument.
Let $\mathcal{U}^*$ denote the set of MEU functions satisfying Symmetry, OI and the following property which we call w-Dominance: for all $n \geq 1$ and pairwise orthogonal $f, g, f_1, f_2, ..., f_n$ in $\mathcal{F}_{\text{fin}}$ such that $f$ dominates $g$,

$$\frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^o \geq \frac{1}{2} f \cdot F_n^o + \frac{1}{2} g \cdot F_n^e,$$

where

$$F_n^e = \frac{1}{2^{n-1}} \left(1 + \sum_{\emptyset \neq I \subset \{1, ..., n\}} \prod_{i \in I} f_i\right) \quad \text{and} \quad F_n^o = \frac{1}{2^{n-1}} \left(\sum_{\emptyset \neq I \subset \{1, ..., n\}} \prod_{i \in I} f_i\right).$$

Note that $F_n^e - F_n^o = \frac{1}{2^{n-1}} \prod_{i=1}^n (1 - f_i) \geq 0$. Take $k \geq 0$ so that $T^k F_n^o$ is orthogonal to $f, g$ and $F_n^e$. Then, by Dominance,

$$\frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot T^k F_n^o \geq \frac{1}{2} f \cdot T^k F_n^o + \frac{1}{2} g \cdot F_n^e.$$

By (B.1) and Symmetry, this is equivalent to (B.2). Hence Dominance implies w-Dominance. Therefore, it suffices to show that any element in $\mathcal{U}^*$ has the required representation.

It is not difficult to show that $\mathcal{U}^*$ is convex. We wish to show that each extreme point of $\mathcal{U}^*$ lies in $\mathcal{V}$, since then we can apply the Choquet Theorem (Phelps (2001, p. 14) to get the representation. The proof re extreme points proceeds as in Proposition B.7 of ES. We show that for any $f^* \in \mathcal{F}_{\{1, ..., m\}}$ and $U \in \mathcal{U}^*$, the two utility functions on $\mathcal{F}$ defined by

$$U^*(f) = \frac{U(f^* \cdot T^m f)}{U(f^*)} \quad \text{and} \quad U^{**}(f) = \frac{U(f) - U(f^* \cdot T^m f)}{1 - U(f^*)},$$

also lie in $\mathcal{U}^*$. The only nontrivial new argument is to show that $U^*$ and $U^{**}$ satisfy w-Dominance, which we do next. This will complete the proof.

$U^*$ satisfies w-Dominance: Take $f, g, f_1, ..., f_n$ as in the statement of w-Dominance. To simplify notation, suppose that $f^*$ is orthogonal to each of these acts. Recall that $f = \pi_1(\alpha h + (1 - \alpha) h')$ and $g = \pi_2(\alpha h + (1 - \alpha) T^k h')$, for some $\pi_1, \pi_2 \in \Pi$ and $h, h' \in \mathcal{F}_{\{1, ..., k\}}$. Then $f^\#$ dominates $g^\#$, where $f^\# \equiv f^* \cdot f$ and $g^\# \equiv (T^{K_1} f^*) \cdot \alpha \pi_2 h + (T^{K_2} f^*) \cdot \alpha \pi_2 h'$, and where $K_1$ and $K_2$ are large enough.
so that $T^K_1 f^*$ is orthogonal to $f^*, T^K_2 f^*, f, g, f_1, ..., f_n$. Therefore, because $U$ satisfies w-Dominance,

$$U \left( \frac{1}{2} f^* \cdot F_n^e + \frac{1}{2} g^* \cdot F_n^o \right) \geq U \left( \frac{1}{2} f^* \cdot F_n^e + \frac{1}{2} g^* \cdot F_n^e \right).$$

By OI,

$$U^* \left( \frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^o \right) = U \left( \frac{1}{2} f^* \cdot F_n^e + \frac{1}{2} g^* \cdot F_n^o \right) \geq U \left( \frac{1}{2} f^* \cdot F_n^e + \frac{1}{2} g^* \cdot F_n^e \right)$$

$$= U^* \left( \frac{1}{2} f \cdot F_n^o + \frac{1}{2} g \cdot F_n^e \right),$$

which proves w-Dominance for $U^*$.

$U^*$ satisfies w-Dominance: Take $f, g, f_1, ..., f_n$ and $f^*$ as above. Note that

$$U^* \left( \frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^o \right) \geq U^* \left( \frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^e \right)$$

is equivalent to

$$\frac{1}{2} U \left( \frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^o \right) - \frac{1}{2} U \left( f^* \cdot \left( \frac{1}{2} f \cdot F_n^e + \frac{1}{2} g \cdot F_n^o \right) \right)$$

$$\geq \frac{1}{2} U \left( \frac{1}{2} f \cdot F_n^o + \frac{1}{2} g \cdot F_n^e \right) - \frac{1}{2} U \left( f^* \cdot \left( \frac{1}{2} f \cdot F_n^o + \frac{1}{2} g \cdot F_n^e \right) \right).$$

Use (B.1) and rearrange the latter inequality to get

$$U \left( \frac{1}{2} f \cdot \left[ \frac{1}{2} F_n^e + \frac{1}{2} f^* \cdot F_n^o \right] + \frac{1}{2} g \cdot \left[ \frac{1}{2} F_n^o + \frac{1}{2} f^* \cdot F_n^e \right] \right)$$

$$\geq U \left( \frac{1}{2} f \cdot \left[ \frac{1}{2} F_n^o + \frac{1}{2} f^* \cdot F_n^e \right] + \frac{1}{2} g \cdot \left[ \frac{1}{2} F_n^e + \frac{1}{2} f^* \cdot F_n^o \right] \right).$$

Letting $f_{n+1} = f^*$ gives $\frac{1}{2} F_n^e + \frac{1}{2} f^* \cdot F_n^o = F_{n+1}^e$ and $\frac{1}{2} F_n^o + \frac{1}{2} f^* \cdot F_n^e = F_{n+1}^o$. Thus, the inequality we need is

$$U \left( \frac{1}{2} f \cdot F_{n+1}^e + \frac{1}{2} g \cdot F_{n+1}^o \right) \geq U \left( \frac{1}{2} f \cdot F_{n+1}^o + \frac{1}{2} g \cdot F_{n+1}^e \right),$$

which is guaranteed by w-Dominance for $U$. 

\[ \Box \]

C. Uniqueness: Proof of Theorem 3.1

Recall that $\mathcal{V}$ is compact metric. We consider a dual pair $(C(\mathcal{V}), ca(\mathcal{V}))$, the set of continuous functions and the set of Borel countably additive signed measures. In this proof, equip the two spaces with the weak topologies $\sigma(C(\mathcal{V}), ca(\mathcal{V}))$ and $\sigma(ca(\mathcal{V}), C(\mathcal{V}))$. Note that $\sigma(ca(\mathcal{V}), C(\mathcal{V}))$ induces the weak-convergence topology on $\Delta(\mathcal{V})$. 

53
Each finitely-based act \( f \) induces the continuous map \( \hat{f} : \mathcal{V} \to [0, 1] \), given by

\[
\hat{f}(V) = V(f).
\]

Let \( \hat{F}_{\text{fin}} \) be the set of all such maps. Then \( \hat{F}_{\text{fin}} \) is a convex subset of \( C(\mathcal{V}) \). For details supporting these assertions, see Appendix B.4 of ES.

Let \( G = \{ \alpha \phi \in \hat{F}_{\text{fin}} : \alpha \geq 0, \phi \in \hat{F}_{\text{fin}} \} \), a convex cone. Then \( \mathcal{M} \subset \Delta(\mathcal{V}) \) is comprehensive if and only if: for every \( \mu' \in \Delta(\mathcal{V}) \),

\[
[\mu \in \mathcal{M} \text{ and } \int \phi d\mu' \geq \int \phi d\mu \text{ for all } \phi \in G] \implies \mu' \in \mathcal{M}.
\]

**Lemma C.1.** Suppose that \( \mathcal{M} \subset \Delta(\mathcal{V}) \) is non-empty, weak-convergence closed, convex and comprehensive, and \( \mu^* \in \Delta(\mathcal{V}) \setminus \mathcal{M} \). Then there exists \( \phi \) in the weak closure of \( G \) such that \( \int \phi d\mu^* < \min_{\mu \in \mathcal{M}} \int \phi d\mu \).

**Proof.** Let \( \overline{\mathcal{M}} \subset ca(\mathcal{V}) \) be the smallest subset of \( ca(\mathcal{V}) \) containing \( \mathcal{M} \) and satisfying: for every \( \mu' \in ca(\mathcal{V}) \),

\[
[\mu \in \mathcal{M} \text{ and } \int \phi d\mu' \geq \int \phi d\mu \text{ for all } \phi \in G] \implies \mu' \in \mathcal{M}.
\]

Then \( \overline{\mathcal{M}} \) is a convex cone in \( ca(\mathcal{V}) \), and \( \mu^* \notin \overline{\mathcal{M}} \).

Recall that the dual cone of \( G \) is

\[
G^+ = \{ \nu \in ca(\mathcal{V}) : \int \phi d\nu \geq 0 \text{ for all } \phi \in G \},
\]

and \( G^{++} = \{ \phi \in C(\mathcal{V}) : \int \phi d\nu \geq 0 \text{ for all } \nu \in G^+ \} \).

Step 1. There exists \( \phi' \in C(\mathcal{V}) \) such that \( \int \phi' d\mu^* < \min_{\mu \in \overline{\mathcal{M}}} \int \phi' d\mu \leq \min_{\mu \in \mathcal{M}} \int \phi' d\mu \): The strict inequality follows by the Separating Hyperplane Theorem (Corollary 5.80 in Aliprantis and Border (2006)). The weak inequality is obvious.

Step 2. \( \phi' \in G^{++} \): Suppose \( \mu \in G^+ \). Take \( \bar{\mu} \in \overline{\mathcal{M}} \) and let \( \mu_\alpha = \alpha \mu + \bar{\mu} \) for each \( \alpha > 0 \). Note that \( \mu_\alpha \in \overline{\mathcal{M}} \) because \( \mu \in G^+ \) implies that \( \int \phi d\mu_\alpha \geq \int \phi d\bar{\mu} \) for each \( \phi \in G \). In addition, \( \int \phi' d\mu_\alpha = \alpha \int \phi' d\mu + \int \phi' d\bar{\mu} \). If \( \int \phi' d\mu < 0 \), then Step 1 cannot be true because \( \alpha \) can be arbitrarily large. Thus, \( \int \phi' d\mu \geq 0 \). Since this is true for any \( \mu \in G^+ \), \( \phi' \in G^{++} \).

Complete the proof. By Lemma 3 of Craven and Koliha (1977), \( G^{++} \) equals the weak closure of \( G \). Moreover, \( \phi \mapsto \int \phi d\mu^* - \min_{\mu \in \mathcal{M}} \int \phi d\mu \) is weakly continuous.
by the Maximum Theorem. Thus, \( \phi' \) can be chosen from \( G \), or equivalently from \( \mathcal{F}_{\text{fin}} \).

Suppose that \( \mathcal{M}, \mathcal{M}' \subset \Delta(\mathcal{V}) \) are non-empty, weak-convergence closed, convex and comprehensive, and that

\[
\inf_{\mu \in \mathcal{M}'} \int_{\mathcal{V}} V(f)\,d\mu(V) = \inf_{\mu \in \mathcal{M}} \int_{\mathcal{V}} V(f)\,d\mu(V) \text{ for all } f \in \mathcal{F}_{\text{fin}}.
\]

If \( \mu^* \in \mathcal{M}' \setminus \mathcal{M} \), there is a contradiction with the lemma. Conclude that \( \mathcal{M} = \mathcal{M}' \).

**D. Proof of Theorem 4.1**

(ii) Prove sufficiency of the axiom (necessity is trivial). Take any \( f, g \in \mathcal{F}_{\{1, \ldots, n\}} \), and \( \pi \in \Pi \) such that \( g, f \) and \( \pi f \) are pairwise orthogonal. By indifference to parameter ambiguity,

\[
W(\alpha f + (1 - \alpha) f \cdot g) = \alpha W(f) + (1 - \alpha) W(f \cdot g).
\]

The functional form (3.5) for \( W \) implies that

\[
W(\alpha f + (1 - \alpha) f \cdot g) = W(\alpha \pi f + (1 - \alpha) f \cdot g).
\]

(See the necessity proof in Appendix A.1.) These two equalities imply

\[
W(\alpha \pi f + (1 - \alpha) f \cdot g) = \alpha W(f) + (1 - \alpha) W(f \cdot g),
\]

and hence the axiom called Orthogonal Independence used in the characterization of the single prior model described in Theorem B.1.

Uniqueness is proven in ES (Thm. 5.2). Alternatively, it is implied by Theorem 3.1 and the fact that

\[
\mu_1 \neq \mu_2 \implies \{ \mu' \in \Delta(\mathcal{V}) : \mu' \triangleright \mu_1 \} \neq \{ \mu' \in \Delta(\mathcal{V}) : \mu' \triangleright \mu_2 \}.
\]

(ii) Again prove sufficiency. Take \( f \in \mathcal{F}_{\{1, \ldots, n\}} \) and \( \pi \in \Pi \). Choose \( m \geq n \) such that \( \pi f \in \mathcal{F}_{\{1, \ldots, m\}} \). Then,

\[
\alpha f + (1 - \alpha) \pi f \sim \alpha f + (1 - \alpha) T^m(\pi f) \sim \alpha f + (1 - \alpha) T^m f \\
\sim \alpha f + (1 - \alpha) f = f.
\]
Indifference to ambiguity about heterogeneity is used for the first and third indifference, and the second one follows from Symmetry. Thus $\alpha f + (1 - \alpha) \pi f \sim f$. This is the SE axiom shown in ES (Thm. 3.2) to characterize the single likelihood model.

For uniqueness, let $M, M' \subset \Delta (\mathcal{V}_{SEU})$ each represent preference through (4.2). By Theorem 3.1, their comprehensive hulls are identical. But since all priors here have support in $\mathcal{V}_{SEU}$, we can say more. We claim that $\mu' \succ \mu$ for $\mu', \mu \in \Delta (\mathcal{V}_{SEU})$ implies $\mu' = \mu$. The simple proof follows.

We can identify $\mu'$ and $\mu$ with probability measures on $\Delta (S)$, since each IID utility in $\mathcal{V}_{SEU}$ corresponds to a unique $\ell$ in $\Delta (S)$ as in (3.2). Letting $f$ be the bet on $s$ in the first experiment, $\mu' \succ \mu$ implies that

$$\int \ell (s) \, d\mu' (\ell) \geq \int \ell (s) \, d\mu (\ell)$$

for every $s \in S$. It follows that equality prevails. Similarly, for $n$-fold products,

$$\int \ell^n (\cdot) \, d\mu' (\ell) = \int \ell^n (\cdot) \, d\mu (\ell).$$

It follows that equality holds also for the infinite products. But then $\mu' = \mu$ by uniqueness of the prior in the de Finetti theorem.

Finally, suppose that $\mu' \in M' \setminus M$. By Theorem 3.1, $\mu'$ is contained in the comprehensive hull of $M$. But then $\mu' \succ \mu$ for some $\mu \in M$ and thus $\mu' = \mu \in M$, a contradiction. Conclude that $M = M'$.

References


