

2013s-35

## **De Finetti Meets Ellsberg**

*Larry G. Epstein, Kyoungwon Seo*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal**  
**Septembre 2013**

© 2013 *Larry G. Epstein, Kyoungwon Seo*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.*

### **Les partenaires du CIRANO**

#### **Partenaire majeur**

Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie

#### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Banque Scotia  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Investissements PSP  
Ministère des Finances et de l'Économie  
Power Corporation du Canada  
Rio Tinto Alcan  
State Street Global Advisors  
Transat A.T.  
Ville de Montréal

#### **Partenaires universitaires**

École Polytechnique de Montréal  
École de technologie supérieure (ÉTS)  
HEC Montréal  
Institut national de la recherche scientifique (INRS)  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

ISSN 2292-0838

Partenaire financier

Enseignement supérieur,  
Recherche, Science  
et Technologie  
Québec 

# De Finetti Meets Ellsberg<sup>\*</sup>

*Larry G. Epstein<sup>†</sup>, Kyoungwon Seo<sup>‡</sup>*

## Résumé/abstract

The paper outlines an exchangeable non-Bayesian model of preference generalizing the Savage/de Finetti classic model of subjective expected utility preference with an exchangeable prior. The treatment is informal, and the emphasis is on motivation and potential applications rather than on axiomatic foundations and technical details. The objective is to provide a widely accessible introduction to research that is reported in detail elsewhere.

**Mots clés/Keywords:** Savage/de Finetti classic model, Ellsberg Paradox.

---

<sup>\*</sup> We gratefully acknowledge the financial support of the National Science Foundation (awards SES-0917740, 0918248 and 1216339). Epstein is grateful also for the generous hospitality of CIRANO where some of this work was completed. This paper was written for the International Workshop in Economic Theory, Ca.Foscari University of Venice, June 28-29, 2013.

<sup>†</sup> CIRANO and Boston University, [lepstein@bu.edu](mailto:lepstein@bu.edu).

<sup>‡</sup> Korea Advanced Institute of Science and Technology, [kseo@kaist.ac.kr](mailto:kseo@kaist.ac.kr).

## 1. Introduction

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 for being unable to accommodate a role for ambiguity, or confidence in beliefs, such as illustrated by the Ellsberg Paradox. This critique motivated development of a number of generalizations of SEU; a prominent example is maxmin expected utility theory due to Gilboa and Schmeidler (1989). The popularity of SEU is due in part to its appealing and elegant axiomatization by Savage (1972). However, its prominence, particularly in statistical decision-making and in learning models, is due also in large part to de Finetti (1937) who considered settings with repeated experiments (or random events—coin tossing is a common example) 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; learning is then modeled by Bayesian updating of beliefs about the parameter. This celebrated result is the cornerstone of Bayesian learning theory, and Kreps (1988) refers to it as "the fundamental theorem of (most) statistics" because of the justification it provides for the analyst to view samples as being independent and identically distributed with unknown distribution function.

However, de Finetti's specialization of SEU, typically called the Bayesian exchangeable model, is also subject to criticisms. To describe them introduce some formalism. Consider a countable number of experiments, each of which yields an outcome in the (finite) set  $S$ . Thus  $\Omega = S^\infty$  is the set of all possible sample paths. An individual must choose between bets, or more generally between acts over  $\Omega$ , where an act is a mapping  $f$  from  $\Omega$  into a set of payoffs. She maximizes expected utility subject to her beliefs about the outcomes of experiments, where beliefs are represented by a probability measure  $P$ , called her *predictive prior*.<sup>1</sup> Say that  $P$  is *exchangeable* if the probability of a finite sequence of outcomes does not depend on the order in which they are realized. De Finetti showed that exchangeability is equivalent to the following "conditionally i.i.d." representation: There exists a probability measure  $\mu$  on  $\Delta(S)$ , the set of probability laws on  $S$ , such that

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

---

<sup>1</sup>We distinguish between a predictive prior, which describes beliefs about the outcomes of experiments, and a prior, such as  $\mu$  in (1.1) below, which describes beliefs about unknown parameters.

where, for any probability measure  $\ell$  on  $S$ ,  $\ell^\infty$  denotes the corresponding i.i.d. product measure on  $\Omega$ . Here  $\ell$  is the unknown parameter and  $\mu$  describes prior beliefs about the parameter. Thus  $P$  is an average, (or expectation, or mixture) of i.i.d. measures. Though intuitively appealing and elegant, the representation (1.1) is suggestive of features that might be viewed as overly restrictive in both statistical analysis and in 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. There are other senses in which the Bayesian exchangeable model assumes self-confidence that may strike one as excessive or extreme in complicated environments. Though there is uncertainty about 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. Moreover, assuming Bayesian updating, under well-known conditions the model implies that the individual is certain that she will learn the true probability law  $\ell^*$  asymptotically given enough data (Savage, 1972, Ch. 3.6). Thus she will learn also the law describing the sequence of experiments—the i.i.d. product  $(\ell^*)^\infty$ —which is everything she needs to know in order to make decisions given that experiments are inherently random and only the governing probability law is knowable at best. Accordingly, the theory leaves no room for doubt about what is well understood or for more modest ambitions about what is learnable.

This paper proceeds as follows. First, Section 2 provides economic motivation for the preceding critique through examples taken from the applied literature where we understand researchers to have expressed similar concerns (to varying degrees and possibly in different language). We also consider Ellsberg-style urns as a canonical example that we use subsequently to illustrate and interpret the issues and models. Though the paper is largely informal and the treatment of models is not axiomatic, we nevertheless translate the above critiques into behavioral terms; Section 3 does this in the context of the urns example. This is the decision-theoretic component of the paper: interpretations and arguments about preference specifications are made precise by expressing them behaviorally.

The core of the paper outlines a model that generalizes the Bayesian exchangeable model and that is designed to relax the strong assumptions noted above.<sup>2</sup> Our

---

<sup>2</sup>The model is a convenient special case of the general axiomatic model in Epstein and Seo (2012); related axiomatically founded models appear in (2010, 2013). We borrow also from Epstein and Schneider (2007); for example, our running example of a sequence of Ellsberg urns is adapted from one of their examples. The latter paper is designed for dynamic applications

view is that the above critique is unrelated to the assumption of exchangeability—as will be amply demonstrated, symmetry is often a natural assumption—but stems rather from the assumption in SEU that beliefs can be represented by a probability measure.<sup>3</sup> Accordingly, the model we describe combines the inspirations of Ellsberg and de Finetti in that it adapts to a setting of repeated experiments models of ambiguity averse preferences (specifically, the maxmin model of Gilboa and Schmeidler (1989)) that were designed to accommodate the behavior pointed to by Ellsberg, thus paralleling de Finetti’s adaptation of SEU. Just as de Finetti’s model is generally thought of in prescriptive terms, the model we describe is primarily normative:<sup>4</sup> it is intended to guide choice by sophisticated thoughtful individuals who are aware of the limits of their understanding of a complex environment (these are not boundedly rational or behaviorally biased individuals). Section 4 models *ex ante* choice, Section 5 illustrates its application, and Section 6 considers how to expand the model to include updating and inference. In the latter connection our objective is not to model econometricians *per se*, but rather to help decision-makers who, as part of their decision-making process, may wish to draw inferences from available data in order to lead to more informed choices. Accordingly, we take preference as the primitive and rules for inference are derived from assumptions about preference. It is well known that ambiguity poses difficulties for updating and that there is no consensus updating rule analogous to Bayes’ Rule. The thrust of this section is two-fold. First, we employ Epstein and Seo (2011, Theorem 2.1) to describe the precise modeling trade-off between dynamic consistency in updating, exchangeability (or symmetry) and ambiguity aversion. This gives a unifying perspective on various updating rules that have been proposed, including the dynamically consistent rule in Epstein and Schneider (2003a, 2007), and also suggests alternative directions for further study. Second, we outline an updating rule axiomatized in our (2010) paper that satisfies dynamic consistency in a limited but important class of environments, and that has a number of interesting features: It models learning that is less ambitious—some but not necessarily all unknown features will be learned—and it prescribes inference when signals are difficult to interpret.

Though the paper describes significant progress towards addressing the cri-

---

typical in macroeconomics and finance, while our focus is (and has been) on cross-sectional applications.

<sup>3</sup>Thus the culprit in our view is probabilistic sophistication (Machina and Schmeidler, 1992).

<sup>4</sup>There is a descriptive component to such modeling given the experimental evidence in Epstein and Halevy (2013).

tiques of the exchangeable Bayesian model, a message we would like readers to take away is that there remain interesting projects for future research, including both applications of the generalized exchangeable model to applied questions in IO, for example, and also further theoretical developments, particularly in modeling updating.

## 2. Motivating Examples

### 2.1. Entry games: multiple equilibria

A policy maker (PM) must choose a policy for a number of markets, the consequences of which depend on how firms in these markets behave. Her view is that in the  $i^{\text{th}}$  market,  $i = 1, \dots, I$ , two firms (who differ across markets) play the entry game with payoffs (in utils) given by the following matrix:

	out	in
out	0, 0	0, $1 - \varepsilon_{2i}$
in	$2 - \varepsilon_{1i}, 0$	$\theta_1 - \varepsilon_{1i}, \theta_2 - \varepsilon_{2i}$

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. The set of possible outcomes in each market is  $S = \{0, 1\} \times \{0, 1\}$ , where  $(1, 0)$  indicates that firm 1 chooses in and firm 2 chooses out, and so on. Obviously, an experiment in our earlier language is identified here with a market.<sup>5</sup>

For concreteness adapt the set up in Ciliberto and Tamer (2009) by taking firms 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 (firm 1) and the other is small (firm 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

---

<sup>5</sup>All of the axiomatizations to which we refer, including de Finetti's seminal result, assume infinitely many experiments. That does not require that there literally be infinitely many markets, for example; it is enough that the PM view the actual set of  $I$  markets as a subset of an infinite set where she would satisfy exchangeability and other relevant axioms. Of course, even that is not necessary if one wishes only to apply the functional forms. Accordingly, we assume below in both the model description and in all examples that  $I \leq \infty$ .

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  $I'$  indexes markets that begin or end at the constructed airport.<sup>6</sup>

A key element of PM's theory is that both  $\theta$  and  $\varepsilon_i$  are known to the firms and that they play a mixed strategy Nash equilibrium, simultaneously in all markets. She knows the structure of the game but not the values of  $\theta$  or the  $\varepsilon_i$ 's. She views the latter as distributed i.i.d. across markets according to the measure  $m \in \Delta(\mathcal{E})$ , for some set  $\mathcal{E}$ ; and she has some information about  $\theta$ , perhaps from past experience in other markets, and is able to form a set of priors about its values. However, her remaining uncertainty is more problematic. Importantly, there may be multiple Nash equilibria in any market, and she does not understand at all how selection occurs and thus how it may differ or be related across markets. In other words, ignorance of the selection mechanism creates uncertainty (or ambiguity) about possible heterogeneity and correlation across markets that she cannot describe but that she may suspect is important. As a consequence, she would be unable to make probabilistic predictions of market outcomes even if she knew the parameter  $\theta$ ; when knowledge of all parameters of a theory are inadequate for making probabilistic predictions we refer to the theory as *incomplete*. How should she choose between available policies?

There are data, consisting of outcomes in markets that have previously played out, that can be used to inform choice. Specifically, PM will generally try to infer what these data say about the unknown parameter  $\theta$ , which is assumed to be constant across markets. However, the multiplicity of equilibria and ignorance of selection also make the inference problem nonstandard. That is because they make it difficult to interpret what observed outcomes imply about  $\theta$ . In general, an outcome, say  $(1, 0)$  for example, could mean that  $\theta$  and the realized  $\varepsilon$ s take on values that make  $(1, 0)$  very likely to arise from a unique mixed strategy equilibrium, or it could be that their values are consistent with multiple equilibria and that  $(1, 0)$  is very likely given the selected equilibrium. Given also PM's general unwillingness or inability to assign a probability to  $(1, 0)$  being selected, the question "what is the likelihood of  $(1, 0)$  given  $\theta$ ?" does not have a unique answer, which makes inapplicable inference procedures that come to mind. However inference

---

<sup>6</sup>The example assumes that the size of the new airport does not affect the payoff of the players. In Section 5, we demonstrate how to accommodate also policies that affect payoffs of the airlines.

is conducted, two properties are intuitive. First, it will lead to updated beliefs about  $\theta$  but, because data can be interpreted in more than one way, posteriors do not necessarily converge to certainty about the truth. Second, PM will not learn about selection: Given that she does not understand selection well enough even to theorize about it, she cannot, nor does she expect to, learn about it. Thus even asymptotically with unlimited data, ignorance of selection will persist.

The above example is indicative of the class of complete information entry games studied in the applied IO literature; see, for example, Tamer (2003, 2010), Ciliberto and Tamer (2009) and the references therein. This literature has focussed on identification, estimation and inference with the novel feature being that parameters are only partially identified—for example, in general a (nonsingleton) set of values for  $\theta$  is consistent with a sample even asymptotically. We take from much of that literature the view expressed most forcefully by Manski (2003) and Tamer (2003, 2010), that modelers should avoid assumptions that are driven by convenience (for example, adding an ad hoc assumption about the selection mechanism in order to permit point identification) rather than by economic theory. These authors often have in mind an empirical modeler, but the same principle has merit when applied to a decision maker such as our PM, who is therefore seen as seeking to make decisions that are robust to model uncertainty, say to the i.i.d. assumption across markets.<sup>7</sup>

Conclude this example by noting that it can be generalized in many directions. A broader class of entry games can be considered (Section 5). More significantly, the PM may be unwilling to commit to the hypothesis of Nash equilibrium. For example, she may be confident only that observed play is rationalizable, which makes multiplicity even more prominent. Finally, she may not be certain of the game form and seek robustness also with respect to this uncertainty. For example, Haile and Tamer (2003) point to the free-form nature of most English auctions in practice and advocate making weak assumptions about bidders' behavior. Then equilibrium behavior in each auction is multiple-valued and can be narrowed down and related across auctions only via heroic and often unjustifiable assumptions. Haile and Tamer are concerned with an empirical modeler who is trying to draw inferences about primitives from data on outcomes in a set of related auctions. However, ambiguity about the game form has implications also for an auctioneer who is choosing reserve prices for forthcoming auctions (Aryal and Kim, 2013).

---

<sup>7</sup>The connection between partial identification and axiomatic models of ambiguity averse preference is due to Epstein and Seo (2013).

## 2.2. Cross-country growth and modeling residuals

Suppose that payoffs to an action depend on the realization of multiple random events. For example, let the outcome  $s_i$  of the  $i$ -th experiment be given by an equation of the form

$$s_i = \beta \cdot x_i + \epsilon_i, \quad i = 1, 2, \dots, I. \quad (2.1)$$

Experiments may differ and the vectors  $x_i$  describe the observable heterogeneity. The key issue is the decision-maker's model of the residuals or unobserved heterogeneity  $\epsilon_i$ , which are the source of the uncertainty she faces. If all sources of heterogeneity of which she is aware are included in the  $x_i$ s, then it is natural, as in the exchangeable Bayesian model, that she be indifferent between any two bets on the realization of residuals that differ only in a reordering of experiments. However, the decision-maker may not be confident that the  $x_i$ s describe all relevant differences between experiments, in which case she may not be certain that residuals are identical, or that they are related in any particular way. Though she may not be able to describe further forms of heterogeneity, she may be worried that there are gaps in her understanding that could be important and thus she may wish to take into account their possible existence when making choices.

Brock and Durlauf (2001) emphasize the appropriateness of such a lack of confidence in the context of the cross-country growth literature; an experiment corresponds to a country and the outcome is its growth rate. They point to the open-endedness of growth theories as a reason for skepticism that all possible differences between countries can be accounted for (p. 231), and they emphasize the importance of "heterogeneity uncertainty." Their remarks can be understood to apply both to an econometrician who is trying to identify the sources of differential growth rates, and also to a policy maker who must choose between policies that will affect growth rates.

The model of residuals is also a concern in robust statistics where the  $\epsilon_i$ s are regression residuals and nonrobustness of standard statistical procedures is a motivation (Huber (1981)). This literature focuses primarily on robustness with respect to the prior, and much less so on robustness with respect to the likelihood which will be at least as important in the work we describe. In particular, limited confidence in the assumption that residuals are i.i.d. is not addressed. This is also true in most of the robust Bayesian literature (Kitagawa (2012), for example). However, Menzel (2011) studies a computational method to efficiently perform robust Bayesian inference in games while relaxing the assumption that equilibrium selections are i.i.d. Because it is closest to our model, we discuss it further in

Section 7 after describing our model. None of the preceding literature addresses explicitly the connection to preference and choice.

### 2.3. Urns

As a running example, consider a sequence of  $I$  Ellsberg urns. You are told that each contains 100 balls that are either red or blue. 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 with all draws being simultaneous. You must choose between bets on the outcomes of the sequence of draws.

The ranking of bets depends on how the urns 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 the parameter  $\theta$  would imply a unique (probabilistic) prediction of the outcomes of draws from all urns. However, there is no reason to be confident that the urns are identical except in the isolated case where you are told that they are.

At the other extreme is extreme ignorance, both about the composition of each urn and about how urns are related. 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, consider an intermediate perception. For example, one might perceive that the fraction  $\lambda$  of the 100 balls is selected once and for all by a single administrator and then placed in each urn, while the other  $(1 - \lambda)$  100 vary across urns in a way that is not understood (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 that is common across urns can be thought of as the probability interval for red given by  $J = [\lambda\theta, \lambda\theta + (1 - \lambda)]$ . 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  $J$  is to be

expected. Note that knowledge of  $J$  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. The bottom line is that outcomes of experiments depend on both common factors (parameters), about which prior beliefs may be imprecise, and on factors that vary across experiments in some unknown or poorly understood fashion and that could render experiments heterogeneous.

### 3. Motivating Behavior

Thus far, both in the critique of the exchangeable Bayesian model and in descriptions of the examples, our language has been informal and vague, and we have been referring to unobservables such as "perceptions." Here we describe behavior that is in principle observable and that gives more precise meaning to expressions such as "aversion to ambiguity about how experiments may differ" and so on. In a normative context, one would pose the choices below to the decision-maker in order to elicit whether ambiguity is a concern and subsequently whether the model described in the next section captures it adequately. We describe behavior in the simplest context possible—the sequence of urns example—and leave it to the interested reader to refer to the cited sources for more general treatments. Later we use the behavior identified here as a measuring stick for evaluating models. We note at the outset that the exchangeable Bayesian model is inconsistent with all of them.

The set of outcomes for each draw is  $\{R, B\}$ , and the set of possible sequences of draws is  $\Omega$  given by

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

To emphasize that the draw is from urn  $i$ , we sometimes write  $S_i = \{R_i, B_i\}$ . Consider the choice between bets on the colors drawn. A bet on drawing red from urn  $i$  is also denoted  $R_i$ ; the interpretation is that the individual receives 1 util if red is drawn and 0 utils otherwise. Payoffs are denominated in utils as in game theory.<sup>8</sup> Similarly,  $R_i B_j$  denotes both the obvious event and the corresponding

---

<sup>8</sup>Justification in our context can be provided by adopting an Anscombe and Aumann domain, where it is assumed that: (i) the individual can rank bets that deliver state-contingent lotteries

bet. Given any two (events and) bets  $A', A \subset \Omega$ , we consider also their mixture  $\frac{1}{2}A' + \frac{1}{2}A$ , which is the bet (or in more common usage, the act) that yields the following payoffs:

$$\frac{1}{2}A' + \frac{1}{2}A \equiv \begin{bmatrix} 1 & A' \cap A \\ 0 & \Omega \setminus (A' \cup A) \\ 1/2 & A' \setminus A \\ 1/2 & A \setminus A' \end{bmatrix}.$$

The interpretation is that after all draws have been completed, an unbiased coin is tossed to determine whether one receives the payoff determined by  $A'$  or that determined by  $A$ . This randomization induces a lottery with prizes denominated in utils. Therefore, if vNM theory is applied to lotteries, one can simply average payoffs as indicated. Randomizing between bets thus smooths out some of the uncertainty associated with the individual bets. Gilboa and Schmeidler (1989) adapt finance terminology and refer to uncertainty being *hedged* through randomization.

We suggest several intuitive properties of preference over bets and their mixtures. Each is interpretable as expressing aversion to a particular form of ambiguity.

Consider first bets on the color of the ball drawn from the first urn. Because there is no information about the composition of the urn, an individual might value randomization and exhibit the ranking<sup>9</sup>

$$\frac{1}{2}R_1 + \frac{1}{2}B_1 \succ R_1 \sim B_1. \quad (3.1)$$

This is the intuitive "ambiguity averse" behavior pointed to by Ellsberg in his two-urn example, as expressed by Gilboa and Schmeidler (1989). The mixture hedges all uncertainty and 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. The remaining rankings reflect the perceived relation between urns and thus relate specifically to repeated experiments.

A more novel ranking to consider is

$$\frac{1}{2}R_1 + \frac{1}{2}B_1 \succ \frac{1}{2}R_1 + \frac{1}{2}B_2. \quad (3.2)$$

---

as payoffs; and (ii) lotteries (acts that are constant over states) are ranked by von Neumann-Morgenstern theory. From (ii), one can infer a vNM index that can be used to measure util payoffs.

<sup>9</sup>The indifference  $R_1 \sim B_1$  is intuitive in the absence of any information that distinguishes between colors.

If urns are thought to be identical, there is no meaningful difference between mixing bets on the same urn versus bets on different urns. Thus one would expect indifference. However, if there is (*aversion to*) *ambiguity about heterogeneity*, then the strict preference indicated is intuitive. The mixture on the left eliminates all ambiguity, (indeed, its payoff is  $\frac{1}{2}$  with certainty), while the mixture on the right moderates ambiguity about factors that are common across experiments, that is, about parameters, but it does not hedge ambiguity about how urns may differ.

We emphasize that even if urns are not viewed as identical, they are naturally taken to be indistinguishable in the sense of indifference between any two bets on sequences of draws that differ only in the numbering of urns—this is the preference version of de Finetti’s exchangeability property for probability measures. The conceptual distinction between indistinguishable and identical experiments is due to Epstein and Schneider (2003b). It reflects the difference between "symmetry of evidence" and "evidence of symmetry" (Epstein and Seo, 2010).

The following ranking is also interpretable in terms of a concern about heterogeneity:

$$\frac{1}{2}B_1R_2 + \frac{1}{2}R_1B_2 \succ R_1B_2 \sim B_1R_2. \quad (3.3)$$

The latter indifference says that urns are indistinguishable (or exchangeable). The strict ranking is the main point. A good scenario for  $B_1R_2$  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  $R_1B_2$ . Thus  $\frac{1}{2}B_1R_2 + \frac{1}{2}R_1B_2$  smooths ambiguity about differences, which motivates (3.3).

Finally, consider

$$\frac{1}{2}R_1 + \frac{1}{2}B_2 \succ R_1 \sim B_2. \quad (3.4)$$

Information about the urns gives no reason for distinguishing between bets on the same color in different urns, or between bets on different colors in the same urn. Therefore, it is intuitive (even compelling) that  $R_1 \sim R_2 \sim B_2$ , and transitivity implies the indifference in (3.4). Turn to the strict preference. 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, the interval  $J = [\lambda\theta, \lambda\theta + (1 - \lambda)]$  in the perception of urns described earlier), and if there is ambiguity about the correct interval. Then mixing between bets on red and blue is valuable, as in the classic Ellsberg experiment, even where the bets are on the draws from different urns. Thus we interpret (3.4) as indicating *aversion to ambiguity about parameters*.

The exchangeable Bayesian model contradicts each of the behaviors (3.1)-(3.4). Next we describe a non-Bayesian model that accommodates them all.

## 4. A Non-Bayesian Exchangeable Model

### 4.1. Utility

There are  $I \leq \infty$  experiments, indexed  $i = 1, \dots, I$ . The ordering of experiments is not temporal, nor is it important. One should think of a cross-sectional setup, where experiments are conducted simultaneously. It is convenient to fix an order, which we do, but it is arbitrary. Each experiment yields an outcome in the finite set  $S$ . The payoff to any chosen physical action depends on the entire sequence of outcomes and thus on the realized state in the state space  $\Omega$  given by

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

The decision-maker chooses between acts, where each act  $f$  is a function from  $\Omega$  into real-valued payoffs denominated in utils.<sup>10</sup> The set of all acts is  $\mathcal{F}$ . To describe choice between acts, we adopt as a primitive the preference  $\succsim$  on  $\mathcal{F}$ . In this section we describe  $\succsim$  through its utility function  $U : \mathcal{F} \rightarrow \mathbb{R}$ . It is to be thought of as describing preference conditional on a fixed and suppressed sample of outcomes for related experiments; ex ante preference corresponds to the case where the sample is empty. Dependence on the sample, that is, inference, is treated explicitly in Section 6.

The benchmark is de Finetti's model, which we rewrite more fully here in terms of utility:

$$U(f) = \int_{\Omega} f dP, \text{ and } P(\cdot) = \int_{\Delta(S)} \ell^I(\cdot) d\mu(\ell). \quad (4.1)$$

Following Gilboa and Schmeidler (1989), we generalize by allowing the individual to entertain a nonsingleton set  $\mathcal{P}$  of predictive priors. The conditionally i.i.d. representation suggests that the multiplicity of predictive priors can arise in two ways: (i) the single prior  $\mu$  is replaced by a set  $\mathcal{M}$  of priors; and (ii) the (uncertain) single likelihood is replaced by a (closed and convex) set  $\mathcal{L}$  of likelihoods.<sup>11</sup> More

<sup>10</sup>Technical details are completely ignored.

<sup>11</sup>Troughout each  $\mathcal{L}$  lies in  $\mathcal{K}(\Delta(S))$ , the set of all closed and convex subsets of the probability simplex  $\Delta(S)$ . Priors  $\mu$  over these sets are taken to be defined on a (suitably defined) Borel sigma algebra of  $\mathcal{K}(\Delta(S))$ . Finally, when  $I = \infty$  one has to replace  $\mathcal{L}^\infty$  by its closure in order to ensure existence of minima; alternatively, one could use infima.

precisely, for any  $\mathcal{L} \subset \Delta(S)$ , define

$$\mathcal{L}^I = \{\ell_1 \otimes \dots \otimes \ell_I : \ell_i \in \mathcal{L}\}, \quad (4.2)$$

the set of all product measures that can be constructed by taking selections from  $\mathcal{L}$ . Then we adopt the following functional form for  $\mathcal{P}$  and the corresponding utility function:<sup>12</sup>

$$U(f) = \min_{P \in \mathcal{P}} \int_{\Omega} f dP \text{ and } \mathcal{P} = \left\{ \int \mathcal{L}^I d\mu(\mathcal{L}) : \mu \in \mathcal{M} \right\}. \quad (4.3)$$

De Finetti's model is the special case where  $\mathcal{M} = \{\mu\}$  is a singleton and where  $\mu$  attaches positive probability only to sets  $\mathcal{L}$  that are singletons.

It is useful to write utility in the form

$$U(f) = \min_{\mu \in \mathcal{M}} \int V_{\mathcal{L}}(f) d\mu(\mathcal{L}), \quad (4.4)$$

where

$$V_{\mathcal{L}}(f) = \left( \min_{P \in \mathcal{L}^I} \int_{\Omega} f dP \right). \quad (4.5)$$

Each utility function  $V_{\mathcal{L}}$  is computed as though there is certainty about the set of likelihoods. The Bayesian special case has i.i.d. beliefs. For this reason (and others that we won't give here), refer to each  $V_{\mathcal{L}}$  as an I.I.D. utility function.<sup>13</sup> It can be thought of as the special case of  $U$  for which  $\mathcal{M} = \{\mu\}$  where  $\mu(\mathcal{L}) = 1$  for some set  $\mathcal{L}$ , and thus<sup>14</sup>

$$\mathcal{P} = \mathcal{L}^I. \quad (4.6)$$

For such a utility function  $V_{\mathcal{L}}$ , the interpretation is that each experiment is believed to be driven by some probability law in  $\mathcal{L}$  where  $\mathcal{L}$ , but not necessarily the probability law, is common to all experiments. Put another way,  $\mathcal{L}^I$  contains many nonidentical product measures and different measures from  $\mathcal{L}$  may apply to different experiments, which suggests that heterogeneity is admitted as a concern.

---

<sup>12</sup>If  $\mu$  has finite support  $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$  and  $\mu_n = \mu(\mathcal{L}_n)$ , then

$$\mathcal{P} = \sum_{n=1}^N \mu_n \cdot (\mathcal{L}_n)^I \equiv \{\sum_{n=1}^N \mu_n P_n : P_n \in (\mathcal{L}_n)^I\}.$$

More generally, integration is in the sense of Aumann.

<sup>13</sup>We reserve the lower case acronym i.i.d. for probability measures.

<sup>14</sup>This special case can be found in Walley and Fine (1982).

Next, just as the de Finetti representation generalizes i.i.d. to exchangeability, admit uncertainty about which set  $\mathcal{L}$  applies. If that uncertainty can be represented by a prior  $\mu$ , then  $\mathcal{P}$  is an "average." If there is ambiguity about the correct set  $\mathcal{L}$ , then the expression for  $\mathcal{P}$  in (4.3) is obtained. Being concerned that the outcomes of experiments may not be i.i.d. and having limited prior information about the factor common to all experiments, the individual seeks to make decisions that are robust in both dimensions.

The role of the multiplicity of likelihoods in the form of the sets  $\mathcal{L}$  merits emphasis. Because of de Finetti, we are accustomed to thinking of experiments (or random events) as being parametrized by a probability law, or by a probability level when the experiment is binary. Here they are parametrized instead by a set of probability laws, or by a probability interval in the binary case. One should think of there being ignorance within each set  $\mathcal{L}$ —there is no distribution within  $\mathcal{L}$ . As a result, unlike in the Bayesian model, certainty about the parameter still leaves scope for some ignorance about the state space (as reflected by maxmin utility with  $\mathcal{P} = \mathcal{L}^I$ ).

Consider also two special cases of the model. Refer to the *single likelihood* model if each  $\mu$  in  $\mathcal{M}$  attaches positive probability only to sets  $\mathcal{L}$  that are singletons. Then, in the obvious notation, we can express the corresponding set of predictive priors in the form

$$\mathcal{P} = \left\{ \int_{\Delta(S)} \ell^I d\mu(\ell) : \mu \in \mathcal{M} \right\}. \quad (4.7)$$

The representation suggests ambiguity about the true probability law for each single experiment, but certainty that experiments are i.i.d. Refer to the *single prior* model if  $\mathcal{M} = \{\mu\}$  is a singleton, in which case (4.3) simplifies to

$$\mathcal{P} = \int \mathcal{L}^I d\mu(\mathcal{L}). \quad (4.8)$$

The de Finetti model equals the intersection of these two special cases.

Though the functional forms suggest interpretations, these call for behavioral justification. First note that the model satisfies the natural symmetry condition, thus justifying the descriptor "exchangeable." To elaborate, let  $\pi$  be any finite permutation of  $\{1, 2, \dots, I\}$ . Given an act  $f$ , define the permuted act  $\pi f$  by  $(\pi f)(s_1, \dots, s_I) = f(s_{\pi(1)}, \dots, s_{\pi(I)})$ . Then

$$U(\pi f) = U(f). \quad (4.9)$$

The other implied behaviors we consider are those given in (3.1)-(3.4) for the urns example. Thus the next section describes how the model works in that example.

**Remark 1.** *The above model is intended to give a flavor of the models that have been developed. We mention some variations or generalizations. In the single prior model as defined above, maxmin is used to aggregate over the set of priors in (4.7), but one could adapt other forms of aggregation from the literature on ambiguity averse preferences. See Al Najjar and de Castro (2010), Klibanoff, Mukerji and Seo (2012), and Cerreia-Voglio et al (2013), for example. However, none of these generalizations affects the observation below that a model with single likelihoods cannot rationalize (3.2) or (3.3). There is also an alternative way to model multiple likelihoods (Epstein and Seo, 2013). Roughly, it differs from (4.8) by expanding the set  $\mathcal{L}^I$  of product measures to include also some measures that are not products, thereby modeling ambiguity about the correlation between experiments.*

## 4.2. The Urns Example Revisited

Take  $S = \{R, B\}$ . The binary case permits a simplified description. Each set  $\mathcal{L}$  can be identified with a unique closed interval  $J$  of probabilities for red:

$$J = [\min_{\ell \in \mathcal{L}} \ell(R), \max_{\ell \in \mathcal{L}} \ell(R)] \subset [0, 1],$$

Therefore, utility is defined given a prior  $\mu$ , or a set of priors  $\mathcal{M}$ , over the set of all subintervals of  $[0, 1]$ .

Any of the three alternative perceptions described in Section 2.3 can be accommodated. Certainty that the urns are identical is modeled by the single likelihood special case. If also there is a single prior,  $\mathcal{M} = \{\mu\}$ , then there is no ambiguity about the common composition and one obtains the exchangeable Bayesian model. Alternatively, complete ignorance is modeled by certainty that the interval  $J = [0, 1]$  applies; one obtains an I.I.D. utility function corresponding to (4.6) and  $\mathcal{L} = \Delta(\{R, B\})$ . Finally, the intermediate perceptions described in Section 2.3 can be accommodated by suitable specifications of the set of priors  $\mathcal{M}$ .

Excluding knife-edge specifications, the non-Bayesian model delivers all of the rankings (3.1)-(3.4) described in the last section, which fact comprises the normative case for the model provided herein.<sup>15</sup> In addition, these rankings discriminate

---

<sup>15</sup>A more complete normative argument would demonstrate that the intuitive behavior also

between the single priors and single likelihoods special cases and thus make the distinction between them empirically meaningful.

We give some details supporting these claims. The expression (4.4) for utility indicates that  $U(f)$  is a suitable aggregate of I.I.D. utilities. Therefore, it is enough (with one exception below) to understand what I.I.D. utility functions imply for the rankings of interest. For bets on a single urn, as in (3.1),  $V_{\mathcal{L}}$  accommodates Ellsbergian ambiguity averse behavior exactly as shown by Gilboa and Schmeidler (1989). The second ranking, which amounts to  $\frac{1}{2} \succ \frac{1}{2}R_1 + \frac{1}{2}B_2$ , is satisfied because utility is evaluated via the worst-case scenario and because the functional form for utility allows urns to differ. In particular, the mixture  $\frac{1}{2}R_1 + \frac{1}{2}B_2$  is evaluated as if the first (second) urn has the smallest (largest) proportion for red consistent with the probability interval corresponding to  $\mathcal{L}$ . Here are some details:

$$\begin{aligned}
 V_{\mathcal{L}}\left(\frac{1}{2}R_1 + \frac{1}{2}B_2\right) &= \min_{\ell_1, \ell_2 \in \mathcal{L}} \left(\frac{1}{2}\ell_1(R_1) + \frac{1}{2}\ell_2(B_2)\right) \\
 &= \frac{1}{2} \min_{\ell_1 \in \mathcal{L}} \ell_1(R_1) + \frac{1}{2} \min_{\ell_2 \in \mathcal{L}} \ell_2(B_2) \\
 &= \frac{1}{2} \min_{\ell_1 \in \mathcal{L}} \ell_1(R_1) + \frac{1}{2} \min_{\ell_1 \in \mathcal{L}} \ell_1(B_1) \\
 &= \frac{1}{2} - \frac{1}{2} \left( \max_{\ell_1 \in \mathcal{L}} \ell_1(R_1) - \min_{\ell_1 \in \mathcal{L}} \ell_1(R_1) \right),
 \end{aligned}$$

which is strictly less than  $\frac{1}{2}$  unless  $\mathcal{L}$  is a singleton. The explanation of why (3.3) is accommodated is similar. The worst-case for  $R_1B_2$  is that the first urn is biased towards blue and the second towards red; the opposite pattern is also possible and is given equal weight ( $R_1B_2$  and  $B_1R_2$  are indifferent by exchangeability), but the mixture hedges this uncertainty and thus is preferable.

Finally, for (3.4), it is not enough to focus on I.I.D. utility functions, and

---

*implies* the model. Such a characterization of the model would require an axiomatic treatment that is beyond the scope of this paper. See Section 4.3 for more on the normative case for the model.

therefore we offer the following elementary proof:

$$\begin{aligned}
U\left(\frac{1}{2}R_1 + \frac{1}{2}B_2\right) &= \min_{\mu \in \mathcal{M}} \int V_{\mathcal{L}}\left(\frac{1}{2}R_1 + \frac{1}{2}B_2\right) d\mu \\
&= \min_{\mu \in \mathcal{M}} \int \left(\frac{1}{2}V_{\mathcal{L}}(R_1) + \frac{1}{2}V_{\mathcal{L}}(B_2)\right) d\mu \\
&\geq \frac{1}{2} \min_{\mu \in \mathcal{M}} \int V_{\mathcal{L}}(R_1) d\mu + \frac{1}{2} \min_{\mu \in \mathcal{M}} \int V_{\mathcal{L}}(B_2) d\mu \quad (4.10) \\
&= \frac{1}{2}U(R_1) + \frac{1}{2}U(B_2) \\
&= \frac{1}{2}U(R_1) + \frac{1}{2}U(B_1) = U(R_1).
\end{aligned}$$

It remains to see when the weak inequality is actually strict and also  $R_1 \sim B_1$ . To illustrate how both conditions may arise, suppose that only two probability intervals for red are thought relevant— $I = [\frac{7}{16}, \frac{11}{16}]$  and its reflection about  $\frac{1}{2}$ ,  $I' = [\frac{5}{16}, \frac{9}{16}]$ . (Formally, they correspond to sets  $\mathcal{L}$  and  $\mathcal{L}'$  respectively; less formally, they are the two values thought possible for the parameter that is common to all urns.) Roughly, every urn is seen as being consistent with a color composition that is either biased towards red (in the case of  $I$ ), or towards blue (in the case of  $I'$ ). For simplicity, suppose further that  $\mathcal{M} = \{\mu, \mu'\}$ , where  $\mu(\{I, I'\}) = \mu'(\{I, I'\}) = 1$ , and

$$\mu(I) = \mu'(I').$$

Thus there is ambiguity about the intervals, but the ambiguity is symmetric in colors: it is easily seen that  $R_1 \sim B_1$ . Furthermore, if  $\mu(I) > \frac{1}{2}$ , then the weak inequality above is strict because the worst-case prior for  $R_1$  is  $\mu'$  and that for  $B_2$  (or equivalently for  $B_1$ ) is  $\mu$ , and  $\mu \neq \mu'$ .

The two special cases of the model perform differently with regard to (3.1)-(3.4). Both can accommodate the standard Ellsberg experiment (3.1). However, of the remaining behaviors, the single likelihood model can rationalize only (3.4) and the single prior model can rationalize only (3.2) and (3.3).<sup>16</sup>

### 4.3. Prior beliefs and the LLN

It seems like a daunting task to form priors over sets  $\mathcal{L}$  or equivalently when the experiment is binary, over corresponding probability intervals. A normative exercise should assist the decision-maker to form priors. In the Bayesian context, guidance

---

<sup>16</sup>For the single prior model, the weak inequality in (4.10) is actually an equality. Therefore, the model implies indifference throughout (3.4). Details are left to the reader.

is provided by the classical law of large numbers (LLN) through the connection it affords between beliefs about parameters ( $\ell \in \Delta(S)$ ) and the willingness to pay for various bets on limiting empirical frequencies; the presumption is that the latter is more readily accessible to introspection. A similar result can be established here, through use of a LLN for nonidentical product measures (Hall and Heyde (1980, Theorem 2.19)), for the single prior special case of our model.<sup>17</sup> Naturally, arguments based on a LLN presume infinitely many experiments ( $I = \infty$ ).

Denote by  $\Phi_n(\cdot)(\omega)$  the empirical frequency measure given the sample  $\omega$ ;  $\Phi_n(A)(\omega)$  is the empirical frequency of the event  $A \subset S$  in the first  $n$  experiments. A decision-maker who is certain that experiments are identical and independent maximizes subjective expected utility with an exchangeable predictive prior; denote by  $\mu$  her prior on  $\Delta(S)$ . Then the classical LLN for exchangeable measures implies that, for any  $K$  and events  $A_k \subset S$ ,  $k = 1, \dots, K$ , and for any real numbers  $a_k$ ,<sup>18</sup>

$$\mu(\cap_{k=1}^K \{\ell \in \Delta(S) : \ell(A_k) \geq a_k\}) = U\left(\left\{\omega : \lim_n \Phi_n(A_k)(\omega) \geq a_k \forall k\right\}\right). \quad (4.11)$$

Thus the prior probability assigned to the sets indicated on the left hand side equals the utility, indeed the certainty equivalent, of the right hand side bets on limiting empirical frequencies. Because  $\mu$  is completely determined by its values on the former sets, a decision-maker can calibrate her prior  $\mu$  if she can arrive at certainty equivalents for the indicated bets.

Such a calibration method extends to a decision-maker who satisfies the single prior version of our model. Because she is uncertain about how experiments may differ, she is not certain that empirical frequencies converge. Nevertheless, there exists the following connection between prior beliefs and certainty equivalents for suitable bets on empirical frequencies:

$$\mu(\cap_{k=1}^K \{\mathcal{L} \subset \Delta(S) : \inf_{\ell \in \mathcal{L}} \ell(A_k) \geq a_k\}) = U\left(\left\{\omega : \liminf_n \Phi_n(A_k)(\omega) \geq a_k \forall k\right\}\right). \quad (4.12)$$

For example, imagine the decision-maker in the sequence-of-urns context struggling with the formation of a prior, and in particular with assessing the prior

---

<sup>17</sup>An extension to multiple priors is the subject of current research and will be reported elsewhere. The result we outline here is modeled on a result in Epstein and Seo (2013) which also assumes a single prior but differs in that multiple likelihoods are as described in Remark 1.

<sup>18</sup> $\{\omega : \lim_n \Phi_n(A_k)(\omega) \geq a_k \forall k\}$  denotes both the event and the bet on the event with winning and losing prizes 1 and 0. Similarly below.

probability of the collection of all sets  $\mathcal{L}$  for which the minimum probability of red is at least  $a$ . According to (4.12), that prior probability should equal her certainty equivalent of the bet that, for all  $\delta > 0$ , the empirical frequency of red is at least  $a - \delta$  in all sufficiently large samples.

To obtain a fuller counterpart of the Bayesian calibration result, one needs to consider not only bets on events as in (4.12), but also more general acts, as described in the following theorem. For any  $\omega = (s_1, \dots, s_{i+1}, s_{i+2}, \dots) \in \Omega = S^\infty$ ,  $T^i\omega \equiv (s_{i+1}, s_{i+2}, \dots)$ , the  $i$ -fold shifted sequence. The next theorem, proven in the appendix, provides a way to calibrate a prior.

**Theorem 4.1.** *Let utility correspond to the single prior special case of (4.3). Then, for any  $f_1, \dots, f_K$  acts over  $S_1 = S$ , and for any real numbers  $a_1, \dots, a_K$ ,*

$$\mu \left( \bigcap_{k=1}^K \left\{ \mathcal{L} : \min_{\ell \in \mathcal{L}} \int f_k d\ell \geq a_k \right\} \right) = U \left( \left\{ \omega : \liminf_n \frac{1}{n} \sum_{i=0}^n f_k(T^i\omega) \geq a_k \forall k \right\} \right). \quad (4.13)$$

Moreover, if  $\mu'$  satisfies the same condition for all  $f_k$  and  $a_k$ , then  $\mu' = \mu$ .

## 5. Modeling the Entry Game

Our objective here is to demonstrate, using the entry game example from Section 2, that our model can be applied naturally to concrete settings that have been of interest in the empirical IO literature.

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  $i = 1, \dots, I$ ; denote by  $I$  also the set  $\{1, 2, \dots, I\}$ . For example, payoffs in utils might be constructed by aggregating the payoffs accruing from individual markets. Then any policy is associated with an act  $f$  in  $\mathcal{F}$ . Thus policy choice can be determined by maximizing the utility function  $U$  of the form in (4.3) over the set of feasible acts. It remains to adopt and motivate a particular specification of the set  $\mathcal{P}$  of predictive priors underlying  $U$ .

The presence of multiple equilibria is key. For each  $\theta$  and  $\varepsilon$ , denote the set of equilibrium mixed strategy profiles in any single market by  $\Psi_\theta(\varepsilon)$ . Thus outcomes in any single market are summarized by the equilibrium correspondence  $\Psi_\theta : \mathcal{E} \rightsquigarrow \Delta(S)$ , for each possible  $\theta$ . Because the outcomes in all markets are relevant,

consider the correspondence  $\Psi_\theta^I : \mathcal{E}^I \rightsquigarrow \Delta(S^\infty)$ , where

$$\Psi_\theta^I(\varepsilon_1, \dots, \varepsilon_I) = \bigotimes_{i=1}^I \Psi_\theta(\varepsilon_i) \equiv \{\ell_1 \otimes \dots \otimes \ell_I : \ell_i \in \Psi_\theta(\varepsilon_i) \text{ for all } i\}.$$

Accordingly  $\Psi_\theta^I(\varepsilon_1, \dots, \varepsilon_i, \dots)$  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 randomizations in different markets are stochastically independent. Beliefs on  $\mathcal{E}^I$  are described by the i.i.d. product measure  $m^I$ . Thus the set of possible likelihoods for  $S^I$  (given  $\theta$ ) is  $\mathcal{L}_\theta^I \subset \Delta(S^I)$  given by

$$\mathcal{L}_\theta^I = \int_{\mathcal{E}^I} \Psi_\theta^I(\varepsilon_1, \dots, \varepsilon_I) dm^I,$$

the set of all mixtures, using  $m$  and all possible selections from the sets  $\Psi_\theta(\varepsilon)$ .<sup>19</sup>

If we define

$$\mathcal{L}_\theta = \int_{\mathcal{E}} \Psi_\theta(\varepsilon) dm, \quad (5.1)$$

then

$$\mathcal{L}_\theta^I = \int_{\mathcal{E}^I} \Psi_\theta^I(\varepsilon_1, \dots, \varepsilon_I) dm^I = (\mathcal{L}_\theta)^I, \quad (5.2)$$

where the latter is defined in (4.2). The following special case illustrates:

$$\begin{aligned} &= \int_{\mathcal{E}} \int_{\mathcal{E}} \{\ell_1 \otimes \ell_2 : \ell_i \in \Psi_\theta(\varepsilon_i) \text{ for } i = 1, 2\} dm(\varepsilon_1) dm(\varepsilon_2) \\ &= \int_{\mathcal{E}} \{p \otimes \ell_2 : p \in \mathcal{L}_\theta, \ell_2 \in \Psi_\theta(\varepsilon_2)\} dm(\varepsilon_2) \\ &= \{p \otimes q : p, q \in \mathcal{L}_\theta\}. \end{aligned}$$

Thus the current specification corresponds to the I.I.D. model of beliefs in (4.6).

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  $\int \mathcal{L}_\theta^I d\mu(\theta)$  of predictive priors. Hence, for general  $\mathcal{M}$ , one obtains the following set  $\mathcal{P}$  of predictive

---

<sup>19</sup>The meaning of this notation is as in (4.3). If  $m$  has finite support, then the integral over  $\mathcal{E}^I$  reduces to a finite sum; in general, integration of the random correspondence is in the sense of Aumann.

priors:

$$\mathcal{P} = \bigcup_{\mu \in \mathcal{M}} \left( \int \mathcal{L}_\theta^I d\mu(\theta) \right).$$

The corresponding utility function is defined as in (4.3).

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  $\mathcal{L}_\theta^I$  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.

A number of generalizations are possible. First it is important to recognize that the specification (5.1) corresponds to a very special case of our model. Recall that  $\Psi_\theta(\varepsilon)$  denotes the set of all distributions over outcomes in a single market that can be generated by varying over *all* (mixed strategy) Nash equilibria for the given  $\varepsilon$  and  $\theta$ . Setting each  $\mathcal{L}_\theta$  to be an average over  $\varepsilon$  of these large sets  $\Psi_\theta(\varepsilon)$  builds in complete ignorance about selection and thereby also precludes differences between policy makers in how they regard selection. Neither property is required by the model in Section 4; in particular, the set of priors  $\mathcal{M}$  can vary with the policy-maker, and so also can the sets  $\mathcal{L}$  to which some prior  $\mu$  in  $\mathcal{M}$  assigns positive probability. In particular, for the entry game one can modify the utility specification by taking

$$\bar{\mathcal{L}}_\theta = \int_{\mathcal{E}} \bar{\Psi}_\theta(\varepsilon) dm,$$

where  $\bar{\Psi}_\theta : \mathcal{E} \rightsquigarrow \Delta(S)$ , is a subcorrespondence of  $\Psi_\theta$ , that is, for each  $\varepsilon$  and  $\theta$ ,

$$\bar{\Psi}_\theta(\varepsilon) \subset \Psi_\theta(\varepsilon).$$

At the extreme of no ignorance, each  $\bar{\Psi}_\theta(\varepsilon)$  consists of only a single measure on  $S$ , say  $p_{\theta,\varepsilon}$ , which models a policy-maker who is certain about the probabilistic selection mechanism at work in each market, and thus is able to accommodate multiplicity by the usual rule for reducing multistage lotteries. For each given  $\theta$ , her beliefs are described by  $p_\theta \equiv \int_{\mathcal{E}} p_{\theta,\varepsilon} dm$  for each market and i.i.d. across

markets. In the absence of ambiguity about  $\theta$ , she behaves as in the de Finetti model using the single predictive prior  $\int (p_\theta)^I d\mu(\theta)$ . More generally, she might view  $p_{\theta,\varepsilon}$  as a focal distribution, but not being completely confident in its validity, she might use  $\bar{\Psi}_\theta$  given by

$$\begin{aligned}\bar{\Psi}_\theta(\varepsilon) &= (1-a)\{p_{\theta,\varepsilon}\} + a\Psi_\theta(\varepsilon) \\ &= \{(1-a)p_{\theta,\varepsilon} + aq : q \in \Psi_\theta(\varepsilon)\},\end{aligned}\tag{5.3}$$

where  $a$  lies between 0 and 1. The two extremes of complete ignorance and no ignorance correspond to  $a = 1$  and 0 respectively. Because each set  $\bar{\Psi}_\theta(\varepsilon)$  expands as  $a$  increases, intermediate values for  $a$  provide a simple way to parametrize differences in ambiguity aversion between policy-makers.

Turn to other generalizations. Any finite number of firms is easily accommodated. So is an incomplete information game with Bayesian-Nash equilibria because multiplicity of Bayesian-Nash equilibria generates multiple likelihoods just 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 appended to  $\theta$ .

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:

	out	in	
out	0, 0	$0, \beta_1 x_{1i} - \varepsilon_{2i}$	(5.4)
in	$\beta_2 x_{2i} - \varepsilon_{1i}, 0$	$\beta_1 x_{1i} - \delta_1 - \varepsilon_{1i}, \beta_2 x_{2i} - \delta_2 - \varepsilon_{2i}$	

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  $x_i$ .

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  $\Delta(\{0, 1\} \times \{0, 1\})^X$ . (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  $\Delta(\{0, 1\} \times \{0, 1\})^X$ .) Thus we arrive at the equilibrium correspondence

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

Finally, the set  $\mathcal{P}$  and the utility function  $U$  can be constructed as above.

Just as in the simpler set up discussed earlier, 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 the 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. 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 (5.4) to be the vector  $(z_{ji}, d)$ . Now we can translate the policy of constructing an airport of size  $d$  into 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  $\omega \in S^I$ , 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  $U$  over the corresponding set of acts.

## 6. Updating

For simplicity, take  $I = \infty$  in this section. Another difference adopted here is that we deviate from the pure cross-sectional setting and permit some experiments (those generating the sample) to be resolved earlier than others. We study choice of acts both ex ante and after observing the outcomes  $s^n = (s_1, \dots, s_n)$  of the first  $n$  experiments, where  $n$  is arbitrary.<sup>20</sup> It is helpful to picture an event tree, with generic node  $(n, s^n)$ . To model both choice and inference, we adopt as primitives the set of conditional preferences  $\{\succeq_{n, s^n} : n \geq 0, s^n \in S^n\}$ , where  $\succeq_{n, s^n}$  is the preference on  $\mathcal{F}$  conditional on being at node  $(n, s^n)$ ;  $\succeq_0$ , corresponding to  $n = 0$ , denotes ex ante preference.

A normatively appealing criterion for updating is that the collection of preferences be *dynamically consistent*: For all  $1 \leq n$ , samples  $s^n$ , and acts  $f', f \in \mathcal{F}$ ,

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

and the latter preference is strict if, in addition,  $f' \succ_{n, s^n} f$  for some  $s^n$ .<sup>21</sup> It is well-known that there is a tension between dynamic consistency and ambiguity aversion (or non-Bayesian models more generally). However, it will help to organize our discussion of inference if we first outline a particularly stark result from Epstein and Seo (2011) that describes the modeling trade-offs involved in a setting with repeated exchangeable experiments.

Suppose that: (i)  $\succeq_0$  is exchangeable, that is, it satisfies the appropriate form of (4.9); (ii) unrealized parts of the tree do not matter— $f' \sim_{n, s^n} f$  for any two acts that coincide on the continuation of the tree from node  $(n, s^n)$ —a property commonly called *consequentialism*; (iii) the collection of preferences is dynamically consistent; and (iv) every preference satisfies extremely weak (we would say innocuous) continuity and monotonicity properties. Then  $\succeq_0$  and all conditional preferences are *additive*, that is, they can be represented by utility functions that

---

<sup>20</sup>We remind the reader that experiments have been ordered, though thus far and in the sequel, the specific order does not matter.

<sup>21</sup>Dynamic consistency requires also that conditionals at any node  $(k, s^k)$  be suitably consistent with conditionals at subsequent nodes. For notational simplicity, we have stated the formal property only for  $k = 0$ .

are additive across states. The significance of additivity is that if each conditional preference has the form in Section 4, then each must conform to SEU. Thus it is not possible to extend the model so as to satisfy dynamic consistency and also consequentialism, without the model collapsing back to de Finetti's and thus sharing its limitations. The same applies to all other non-Bayesian exchangeable models that we have cited. One can also point to a behavioral functional-form-free implication of the noted additivity: Given the latter, then aversion to ambiguity about heterogeneity in the sense of (3.2) implies also that<sup>22</sup>

$$\frac{1}{2}R_1 + \frac{1}{2}R_2 \succ R_1. \quad (6.1)$$

A positive value for randomizing between the bets  $R_1$  and  $R_2$  is intuitive in some special scenarios—for example, recalling the perceptions described in Section 2.3, if it is thought that there is only one urn for which the variable component contains a red ball but where the identity of the urn is unknown. However, one would expect indifference in (6.1) if there is no reason for believing that the variable components of urns are related in any particular way. In sum, therefore, for the setting we are modeling, given exchangeability and consequentialism, then dynamic consistency and aversion to ambiguity about heterogeneity imply counterintuitive behavior in the form (6.1).

The lesson is that "one can't have everything." The theorem tells the modeler that if she insists on capturing uncertainty about how experiments may differ, then she must decide which of the three key properties—exchangeability, consequentialism and dynamic consistency—she is willing to relax. This perspective unifies some existing and potential future work. Epstein and Schneider (2003a, 2007) drop exchangeability, which is sensible in their temporal setting, but arguably not when experiments are resolved simultaneously and thus where there is nothing at all, not even time of resolution, that distinguishes between them. They formulate a recursive model which violates exchangeability but satisfies both dynamic consistency and consequentialism. An alternative route that has yet to be explored formally is to relax consequentialism. For example, frequentist statistical procedures employ an *ex ante* perspective whereby all potential data are used in evaluating a procedure rather than just the realized sample, thus violating consequentialism (Berger, 1985). It remains to be seen whether the present concern with robustness and ambiguity leads to specific violations of consequentialism and to specific procedures as being optimal.

---

<sup>22</sup>See Section 2 in Epstein and Seo (2013).

A third route, followed in our papers (2010, 2013), is to relax dynamic consistency, the rationale being that a weaker property is sometimes adequate. Specifically, consider situations where an individual first samples and observes the outcomes of the first  $n$  experiments, and then chooses, once and for all, how to bet on the outcomes of remaining experiments. In particular, the outcomes in markets 1 to  $n$  are ‘pure’ signals and are not payoff relevant, while outcomes in markets  $n + 1$  and beyond influence payoffs but are not a source of information for further updating (which is done only once). In that case, it suffices to have the consistency between ex ante and ex post rankings that is expressed by restricting the above dynamic consistency condition to acts  $f'$  and  $f$  that depend on the outcomes of experiments  $n + 1$ ,  $n + 2$  and so on. Call this weaker property *weak dynamic consistency*.

In the cited papers, we assume also that  $\{\succeq_{n,s^n}\}$  satisfy consequentialism and another property called *commutativity*—each conditional preference  $\succeq_{n,s^n}$  is unchanged if the sample  $(s_1, \dots, s_n)$  is permuted. The rationale is that under the assumption that experiments 1 through  $n$  are resolved simultaneously, that is, data are cross-sectional, then, because there is no natural ordering of cross-sectional data, the order of past observations should not matter. Finally, we assume that every conditional preference conforms to the single priors special case of (4.3).<sup>23</sup> Thus the conditional utility function at  $(n, s^n)$  is given by

$$U_{n,s^n}(f) = \int V_{\mathcal{L}}(f) d\mu_{n,s^n}(\mathcal{L}),$$

where  $V_{\mathcal{L}}(\cdot)$  is defined in (4.5), and where  $\mu_{n,s^n}$  is the posterior belief over the sets  $\mathcal{L}$  of likelihoods. The question at hand is what can be said about how posteriors are derived from the prior  $\mu_0$  for any sample  $s^n$ .

As we have seen, in the urns example, (or more generally for any binary experiments), each set  $\mathcal{L}$  can be identified with a probability interval  $J$  for red, and thus the issue is how to update probabilistic beliefs about intervals. The difficulty is that there is no obvious unique answer to the question "what is the likelihood of observing red given the interval  $J$ ?" Every number in  $J$  would seem to qualify equally as an answer. Based primarily on the preceding assumptions, we show that the individual should form a unique likelihood by averaging over  $J$  and *then use the likelihood constructed in this way and Bayes' rule to update  $\mu_0$* . More precisely, for each  $J$ , she should adopt a probability measure  $\lambda_J$  on  $[0, 1]$ , with

---

<sup>23</sup>Thus ambiguity about parameters is excluded. It is an open question if and how the result outlined below may be extended.

$\lambda_J(J) = 1$ , to construct  $L(R | J)$  according to

$$L(R | J) = \int_{[0,1]} q d\lambda_J(q);$$

and, more generally, for any finite sample  $s^n$ , the constructed likelihood is

$$L(s^n | J) = \int_{[0,1]} q^{\psi_n(s^n)} (1 - q)^{n - \psi_n(s^n)} d\lambda_J(q), \quad (6.2)$$

where  $\psi_n(s^n)$  denotes the number of red balls drawn in the first  $n$  experiments. The measure  $\lambda_J$  can depend on the interval  $J$  and, importantly, also on the individual—because signals are difficult to interpret, individuals may interpret them differently and thus the subjectivity of the measures  $\{\lambda_J\}$  is an appealing feature of the model.

**Remark 2.** *The preceding extends in the obvious way to general (nonbinary) experiments. In that case, a likelihood is constructed by using a measure  $\lambda_{\mathcal{L}}$  on  $\Delta(S)$ ,  $\lambda_{\mathcal{L}}(\mathcal{L}) = 1$ , to average over all i.i.d. measures that can be constructed using measures in  $\mathcal{L}$ . Formally,*

$$L(s^n | \mathcal{L}) = \int_{\Delta(S)} p^\infty(s^n) d\lambda_{\mathcal{L}}(p).$$

*It may be instructive to verify that weak dynamic consistency is implied. Let  $n > 0$  and define*

$$\bar{L}(s^n) = \int L(s^n | \mathcal{L}) d\mu_0(\mathcal{L}).$$

*Then, for any act  $f$  over experiments  $n + 1$  and beyond,*

$$\begin{aligned} \sum_{s^n \in S^n} \bar{L}(s^n) U_{n,s^n}(f) &= \sum_{s^n \in S^n} \bar{L}(s^n) \left( \int V_{\mathcal{L}}(f) d\mu_{n,s^n}(\mathcal{L}) \right) \\ &= \left( \int V_{\mathcal{L}}(f) [\sum_{s^n \in S^n} L(s^n | \mathcal{L})] d\mu_0(\mathcal{L}) \right) \\ &= \int V_{\mathcal{L}}(f) d\mu_0(\mathcal{L}) = U_0(f), \end{aligned}$$

*where the second equality relies on Bayesian updating and on the assumption that the act  $f$  is constant with respect to  $s^n$  (hence the argument does not prove dynamic consistency). Weak dynamic consistency follows if  $\bar{L}(s^n) > 0$  for all  $n$  and  $s^n$ .*

*For greater clarity, we revert to the urns example below.*

A unique likelihood measure on  $\Omega = S^\infty$ , denoted  $L(\cdot | J)$ , is induced by the construction in (6.2) given the  $\lambda_J$ s. Note that this likelihood is exchangeable (and generally not i.i.d.). Exchangeable likelihoods are used also by Acemoglu et al. (2009) to model a difficult to interpret signal. We interpret the likelihood similarly: it is as if the individual is uncertain, in the way represented by  $\lambda_J$ , what any given realized sample reveals about  $J$ . Moon and Schorfheide (2012) use a likelihood function of this form in their Bayesian econometric approach to inference in partially identified models. A possibly surprising feature of our model is that Bayesian inference is compatible with non-Bayesian choice that reflects an aversion to ambiguity about heterogeneity.

The fact that inference is Bayesian-like has the advantage that results from Bayesian learning theory can be adapted to describe the nature of learning in our model. There is a difference in interpretation however. In our model updating takes place only once, after observing  $s^n$  for some given  $n$ . Thus, for example, statements about limiting posteriors describe beliefs if the single updating stage takes place after a very large sample, and not the asymptotics of a dynamic process. With that in mind, consider briefly two important features of learning assuming that each  $\lambda_J$  is the uniform distribution over the interval  $J$  (see our (2010) paper for more details and examples). First, beliefs need not converge to certainty about a specific interval: Where  $\mu_0$  assigns positive probability to overlapping intervals  $J'$  and  $J$ , then along samples  $\omega = (s^n, s_{n+1}, \dots)$  for which the limiting empirical frequency of red  $\lim \psi_n(s^n)/n$  lies in the intersection  $J' \cap J$ , the posterior probability of each interval is positive in general, even in the limit. Second, the sample leads to inferences being drawn about intervals, which constitute the parameters which the individual attempts to learn, but there is no learning within intervals. As described in Section 2.3, intervals with positive length reflect the poorly understood idiosyncratic component of each urn; learning about this component is not attempted.

To illustrate the latter point more concretely, suppose, as in the intermediate perception described in Section 2.3, that the individual believes that  $100\lambda$  balls are selected once and for all by a single administrator and then placed in each urn, while the other  $100(1 - \lambda)$  vary across urns in a way that is not understood. She is certain about the value of  $\lambda$  (for simplicity), but she has a prior  $\mu$  over  $\theta$ , the proportion of red in the common group of balls. Then the relevant probability intervals for red are  $J_\theta = [\lambda\theta, \lambda\theta + (1 - \lambda)]$ ,  $0 \leq \theta \leq 1$ , and  $\mu$  induces a prior over these intervals. The sample is used to draw inferences about  $\theta$ . But the idiosyncratic nature of the  $100(1 - \lambda)$  balls precludes hoping to learn about them—

the color compositions of the  $100(1 - \lambda)$  balls in the sampled urns do not provide any information on the remaining urns. Note that  $\lambda$  is subjective: If  $\lambda = 1$ , intervals are degenerate and the individual views everything as learnable, while if  $\lambda = 0$ , she does not attempt to learn at all.

Similar features are present also when the above updating model is applied to the entry game. Let  $\theta = (\beta_1, \beta_2, \delta_1, \delta_2, \kappa)$  be the parameter vector for the general entry game specification outlined in Section 5. Beliefs about  $\theta$  are updated along a sample, but in general, there will be certainty asymptotically only about a nonsingleton set of values for  $\theta$ , corresponding to the parameter being partially identified; this is due to the fact that multiplicity of equilibria and ignorance of selection imply that the message about parameters inherent in any sample is not unique. In addition, just as in the urns example she views some aspects of her environment as unlearnable, similarly here, the decision-maker has a limited understanding of selection and does not expect her prior ignorance (which is subjective) to be informed by data. When ignorance is modeled as in (5.3),  $a$  captures the size of the unlearnable component. If  $a = 0$ , then selection is described by a single probability law and there is certainty ex ante that all parameters will be learned. If  $a > 0$ , then the individual perceives that to a degree selection is unrelated across markets and therefore unlearnable.

## 7. Concluding Remarks

An overview and summarizing comments may best be expressed through a comparison with Menzel (2011). He also addresses robustness with respect to the i.i.d. assumption, but the focus of his paper differs from ours. As he states (p. 4), his paper "is mainly aiming at providing computational tools," while our primary objective (see particularly our papers dated 2010 and 2013) is to formulate a model that has clear choice-theoretic axiomatic foundations. Note, however, that foundations do not come at the cost of tractability at least for inference, which is Bayesian in our model and hence amenable to familiar computational methods. Another difference in focus is that "choice" in our model refers to "economic choice," where the payoffs to actions depend on the realized outcomes of experiments (or markets), as opposed to Menzel's focus which we would term "statistical choice," where payoffs are assumed to depend on the true values of parameters (such as  $\theta$  in the simple entry game). We emphasize that though economic choice is our primitive, sharp implications for inference are derived, and accordingly, our approach is aptly described as providing a unifying framework

for (economic) choice and inference.

In terms of modeling details, consider the single-prior special case of our model for which we have modeled updating. Then Menzel has in common with us the assumption of a single prior over parameters (such as over  $\theta$  in the simple entry game), and agnosticism about selection modeled through a set of likelihoods (our  $\mathcal{L}$ ).<sup>24</sup> His inference method amounts to applying Bayes' rule for each likelihood function separately and then using the worst-case posterior to evaluate any given statistical action, that is, he minimizes *Gamma-posterior expected loss*. Though this class of statistical decision rules is common in the robust Bayesian literature, to our knowledge it has no decision-theoretic foundations in exchangeable models.<sup>25</sup> There is an apparent formal connection to the Gilboa and Schmeidler (1989) maxmin model; but the latter deals exclusively with a static context and its foundations are at best suggestive of the decision-theoretic meaning of the Gamma-posterior expected loss criterion in a dynamic context that includes updating. In particular, as Menzel acknowledges, it leads to dynamic inconsistency. In contrast, a weak form of dynamic consistency is in large part responsible for the implication in our model that updating should be Bayesian using a likelihood that is constructed by averaging over  $\mathcal{L}$ , or more precisely, over  $\mathcal{L}^I$ . Finally, we note that it is the Gamma-posterior expected loss criterion that creates the tractability/computational issues addressed by Menzel.

## A. Appendix: Proof of Theorem 4.1

For an event  $E \subset \Omega$ , write  $\mathcal{L}^\infty(E)$  instead of  $\inf_{P \in \mathcal{L}^\infty} P(E)$ .

Step 1.  $\mathcal{L}^\infty(\{\omega : \liminf_n \frac{1}{n} \sum_{i=0}^n f(T^i \omega) \geq a\}) = 1$  if  $\min_{\ell \in \mathcal{L}} \int f d\ell \geq a$ : For any  $P = \ell_1 \otimes \ell_2 \otimes \dots \in \mathcal{L}^\infty$ ,

$$\begin{aligned} \liminf_n \frac{1}{n} \sum_{i=0}^n f(T^i \omega) &= \liminf_n \left[ \frac{1}{n} \sum_{i=0}^n f(T^i \omega) - \frac{1}{n} \sum_{i=0}^n \int f d\ell_i + \frac{1}{n} \sum_{i=0}^n \int f d\ell_i \right] \\ &\geq \lim_n \left[ \frac{1}{n} \sum_{i=0}^n f(T^i \omega) - \frac{1}{n} \sum_{i=0}^n \int f d\ell_i \right] + \liminf_n \frac{1}{n} \sum_{i=0}^n \int f d\ell_i \\ &= \liminf_n \frac{1}{n} \sum_{i=0}^n \int f d\ell_i \geq a, P\text{-a.s.} \end{aligned}$$

<sup>24</sup>Menzel assumes complete agnosticism in the sense of  $a = 1$  in (5.3).

<sup>25</sup>Epstein and Schneider (2003a) provide foundations for prior-by-prior Bayesian updating in a temporal recursive model where exchangeability is violated.

The last equality holds by Hall and Heyde (1980, Theorem 2.19).

Step 2.  $\mathcal{L}^\infty \left( \left\{ \omega : \liminf_n \frac{1}{n} \sum_{i=0}^n f(T^i \omega) \geq a \right\} \right) = 0$  if  $\min_{\ell \in \mathcal{L}} \int f d\ell < a$ :  
This follows because  $P \left( \left\{ \omega : \liminf_n \frac{1}{n} \sum_{i=0}^n f(T^i \omega) \geq a \right\} \right) = 0$  if  $P = \ell^\infty$  and  $\int f d\ell < a$ .

Step 3. (4.13) holds: It follows by Steps 1 and 2 because

$$\mathcal{L}^\infty \left( \bigcap_{k=1}^K \left\{ \omega : \liminf_n \frac{1}{n} \sum_{i=0}^n f_k(T^i \omega) \geq a_k \right\} \right) = \begin{cases} 1 & \text{if } \min_{\ell \in \mathcal{L}} \int f_k d\ell \geq a_k \text{ for all } k \\ 0 & \text{otherwise.} \end{cases}$$

Step 4. Uniqueness: Equip  $\mathcal{K}(\Delta(S))$ , the collection of all closed convex subsets of  $\Delta(S)$ , with the Hausdorff metric  $d$ , which renders it compact and separable. Because each  $f_k$  is a function on the finite set  $S$ , it is continuous and, by the Maximum Theorem, so is  $\mathcal{L} \mapsto \min_{\ell \in \mathcal{L}} \int f_k d\ell$ . Therefore, all sets of the form

$$\bigcap_{k=1}^K \left\{ \mathcal{L} : \min_{\ell \in \mathcal{L}} \int f_k d\ell \geq a_k \right\}$$

are closed and lie in  $\Sigma_{\Delta(S)}$ , the Borel  $\sigma$ -algebra of  $\mathcal{K}(\Delta(S))$ . Let  $\Sigma' \subset \Sigma_{\Delta(S)}$  be the  $\sigma$ -algebra generated by sets of the above form.

Claim 1:  $\mu' = \mu$  on  $\Sigma'$ , by Aliprantis and Border (2006, Theorem 10.10).

Claim 2:  $\Sigma_{\Delta(S)} \subset \Sigma'$ . Let  $D$  be a countable dense subset of  $\mathcal{K}(\Delta(S))$  and consider the set  $\mathcal{N}$  of all open balls centered at some point in  $D$  and having a rational radius. Then  $\mathcal{N}$  is a countable basis for the Hausdorff metric topology (Aliprantis and Border, 2006, Lemma 3.4) and  $\mathcal{N}$  generates  $\Sigma_{\Delta(S)}$ . Therefore, it suffices to show that  $\mathcal{N} \subset \Sigma'$ , or that

$$\{\mathcal{L}' : d(\mathcal{L}, \mathcal{L}') < \varepsilon\} \in \Sigma' \tag{A.1}$$

for every  $\varepsilon > 0$  and  $\mathcal{L} \in \mathcal{K}(\Delta(S))$ .

For each  $\mathcal{L}$ , the support function  $\sigma_{\mathcal{L}} : \Phi \rightarrow \mathbb{R}$  (where  $\Phi = \{\phi \in \mathbb{R}^n : \|\phi\| = 1\}$  and  $n = |S|$ ) is defined by  $\sigma_{\mathcal{L}}(\phi) = \min_{\ell \in \mathcal{L}} \int \phi d\ell$ . It is known that

$$d(\mathcal{L}, \mathcal{L}') = \sup_{\phi \in \Phi} |\sigma_{\mathcal{L}}(\phi) - \sigma_{\mathcal{L}'}(\phi)| \equiv \|\sigma_{\mathcal{L}} - \sigma_{\mathcal{L}'}\|$$

(Aliprantis and Border, 2006, Theorem 7.58). Fix  $\varepsilon > 0$  and  $\mathcal{L} \in \mathcal{K}(\Delta S)$ . Since  $\Phi$  is separable, there is a countable set  $\{\phi_1, \phi_2, \dots\}$  that is dense in  $\Phi$ . Let  $a_j = \sigma_{\mathcal{L}}(\phi_j) - \varepsilon$  and  $b_j = \sigma_{\mathcal{L}}(\phi_j) + \varepsilon$ , for each  $j = 1, \dots$ . Then

$$\{\mathcal{L}' : a_j < \sigma_{\mathcal{L}'}(\phi_j) < b_j, \text{ for all } j = 1, \dots\} \in \Sigma'.$$

The noted denseness and the continuity of support functions imply (A.1). ■

## References

- [1] D. Acemoglu, V. Chernozhukov and M. Wernz, Fragility of asymptotic agreement under Bayesian learning, 2009.
- [2] C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis*, 3rd edition, Springer, 2006.
- [3] N. Al Najjar and L. de Castro, Parametric representation of preferences, 2010.
- [4] G. Aryal and D. Kim, A point decision for partially identified auction models, 2013.
- [5] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York, 1985.
- [6] W. Brock and S.N. Durlauf, Growth empirics and reality, *The World Bank Review* 15 (2001), 229-272.
- [7] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Ambiguity and robust statistics, *J. Econ. Theory* 148 (2013), 974-1049.
- [8] F. Ciliberto and E. Tamer, Market structure and multiple equilibria in airline markets, *Econometrica* 77 (2009), 1791-1828.
- [9] B. De Finetti, La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincare* 7 (1937), 1-68. English translation in *Studies in Subjective Probability*, 2nd edition, H.E. Kyburg and H.E. Smokler eds., Krieger Publishing, Huntington NY, 1980, pp. 53-118.
- [10] L.G. Epstein and Y. Halevy, No two experiments are identical, 2013.

- [11] L.G. Epstein and M. Schneider, Recursive multiple-priors, *J. Econ. Theory* 113 (2003a), 1-31.
- [12] L.G. Epstein and M. Schneider, IID: independently and indistinguishably distributed, *J. Econ. Theory* 113 (2003b), 32-50.
- [13] L.G. Epstein and M. Schneider, Learning under ambiguity, *Rev. Ec. Studies* 74 (2007), 1275-1303.
- [14] L.G. Epstein and K. Seo, Symmetry of evidence without evidence of symmetry, *Theor. Econ.* 5 (2010), 313-368.
- [15] L.G. Epstein and K. Seo, Symmetry or dynamic consistency? *B.E. J. Theor. Econ.* (Advances), 2011.
- [16] L.G. Epstein and K. Seo, Ambiguity with repeated experiments, 2012.
- [17] L.G. Epstein and K. Seo, Bayesian inference and non-Bayesian choice and prediction: foundations and an application to entry games with multiple equilibria, 2013.
- [18] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique priors, *J. Math. Econ.* 18 (1989), 141-153.
- [19] P.A. Haile and E. Tamer, Inference with an incomplete model of English auctions, *J. Pol. Econ.* 111 (2003), 1-51.
- [20] P. Hall and C.C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, 1980.
- [21] P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.
- [22] T. Kitagawa, Inference and decision for set identified parameters using posterior lower and upper probabilities, UCL, 2012.
- [23] P. Klibanoff, S. Mukerji and K. Seo, Relevance and symmetry, 2012.
- [24] D.M. Kreps, *Notes on the Theory of Choice*, Westview, 1988.
- [25] M.J. Machina and D. Schmeidler, A robust definition of subjective probability, *Econometrica* 60 (1992), 745-780.

- [26] C.F. Manski, *Partial Identification of Probability Distributions*, Springer, 2003.
- [27] K. Menzel, Robust decisions for incomplete structural models of social interactions, NYU, 2011.
- [28] H.R. Moon and F. Schorfheide, Bayesian and frequentist inference in partially identified models, *Econometrica* 80 (2012), 755-82.
- [29] L.J. Savage, *The Foundations of Statistics*, revised edition, Dover, New York, 1972.
- [30] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003), 147-165.
- [31] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010), 167–195.
- [32] P. Walley and T.L. Fine, Towards a frequentist theory of upper and lower probability, *Ann. Statist.* 10 (1982), 741-761.