A Subjective Foundation of Objective Probability*

Luciano De Castro† *and* Nabil I. Al-Najjar‡

First draft: December 2008; This version: July 2009

Abstract

De Finetti's concept of exchangeability provides a way to formalize the intuitive idea of similarity and its role as guide in decision making. His classic representation theorem states that exchangeable expected utility preferences can be expressed in terms of a subjective beliefs on parameters. De Finetti's representation is inextricably linked to expected utility as it simultaneously identifies the parameters *and* Bayesian beliefs about them. This paper studies the implications of exchangeability assuming that preferences are monotone, transitive and continuous, but otherwise incomplete and/or fail probabilistic sophistication. The central tool in our analysis is a new *subjective ergodic theorem* which takes as primitive preferences, rather than probabilities (as in standard ergodic theory). Using this theorem, we identify the i.i.d. parametrization as sufficient for all preferences in our class. A special case of the result is de Finetti's classic representation. We also prove: (1) a novel derivation of subjective probabilities based on frequencies; (2) a subjective sufficient statistic theorem; and that (3) differences between various decision making paradigms reduce to how they deal with uncertainty about a common set of parameters.

*We thank Larry Epstein, Carlos Moreira, Klaus Nehring, Mallesh Pai, Pablo Schenone, Kyoungwon Seo, Eran Shmaya, Marcelo Viana.

† Department of Economics, University of Illinois, Urbana-Champaign.

‡ Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208.

Contents

1 Introduction

Theories of decision making under uncertainty can be viewed as developing parsimonious representations of the environment decision makers face. A leading example is Savage's subjective expected utility theory. This theory introduces axioms that reduce the problem of ranking potentially complex state-contingent acts to calculating their expected utility with respect to a subjective probability.

A different approach, prevalent in statistics, is to think about inference in terms of "objective parameters" that summarize what is relevant about the states of the world. Parametrizations act as information-compression schemes through which inference and decision making can be expressed on a parsimonious space of parameters, rather than the original primitive states.[1]

De Finetti's notion of exchangeability makes it possible to integrate the subjective and parametric approaches into one elegant theory. Specifically, suppose that an experimental scientist or an econometrician conducts (or passively observes) a sequence of observations in some set S. Since learning from data requires pooling information across experiments, the scientist's or the econometrician's inferences are predicated on the assumption, implicit or explicit, that the experiments are, in a sense, "similar."

De Finetti makes the intuitive idea of similarity formal through his notion of *exchangeability*. Roughly, a decision maker subjectively views a set of experiments as *exchangeable* if he treats the indices interchangeably.[2] Different experiments will usually yield different outcomes, each of which is the result of a multitude of poorly understood factors. Nevertheless, a decision maker's subjective judgment that the experiments are exchangeable amounts to believing that they are governed by the same underlying stochastic structure. De Finetti's celebrated representation says that a probability distribution P

[1]Sims (1996) articulates the view that scientific modeling is a process of finding appropriate data compression schemes via parametric representations.

[2]See de Finetti (1937). Somewhat more formally, exchangeability means that the decision maker ranks as indifferent an act f and the act $f \circ \pi$ that pays f after a finite permutation of the coordinates π is applied. In particular, he considers an outcome s to be just as likely to appear at time t as at time t'.

on Ω is exchangeable if and only if it has the parametric form:

$$P = \int_{\Theta} P^{\theta}\, d\mu(\theta). \tag{1}$$

Here the parameter set Θ indexes the set of all i.i.d. distributions P^{θ} with marginal θ, and μ is a probability distribution on Θ. The decomposition (1) says that a process is exchangeable if and only if it is i.i.d. with unknown parameter.

As it stands, this decomposition is just a mathematical result about probability measures. To link it to decision problems we consider a preference relation $\succcurlyeq$ on acts $f : \Omega \to \mathbb{R}$. The *parameter-based act* $F(\theta) = \int_{\Omega} f\, dP^{\theta}$ expresses f in terms of the parameters, rather than the original states. If $\succcurlyeq$ satisfies Savage's axioms with exchangeable subjective belief P, de Finetti's theorem implies that the decision maker prefers an act f to g if and only if $\int_{\Theta} F\, d\mu \geq \int_{\Theta} G\, d\mu$, where μ is given by (1).

De Finetti's theorem thus integrates subjective beliefs and parametric representations by identifying parameters that are a *sufficient statistic*, in the sense that an exchangeable preference is completely determined by the ranking it induces on parameter-based acts.

⋆ ⋆ ⋆

In de Finetti's classic representation, the identification of parameters is inextricably tied to the expected utility criterion. On the other hand, the concept of parameters as parsimonious representations of what is relevant about a decision problem is meaningful and, indeed, widely used independently of the expected utility criterion. De Finetti's representation as it stands holds little value to, say, classical statistics or the various approaches to model ambiguity. Conceptually, we shall also argue that the subjective belief in the similarity of experiments is of different nature than beliefs over the parameter values.

This paper studies the implications of exchangeability for preferences that are continuous, monotone and transitive, but may otherwise be incomplete and/or fail probabilistic sophistication.[3] These include, as special cases,

[3] In the sense of Machina and Schmeidler (1992).

exchangeable versions of Bewley (1986)'s model of incomplete preferences, Gilboa and Schmeidler (1989)'s multiple prior model, and classical statistics procedures.

Our central result, Theorem 1, is a new *subjective ergodic theorem*. We establish this result for abstract state spaces and transformations; here we illustrate it in the special case of exchangeability. Consider the stylized setting of coin tosses, let f_1 denote the act that pays 1 if the first coin turns *Heads* and 0 otherwise. At a state ω, define $f^\star(\omega)$ as the limiting average payoff of the sequence of acts $f_i, i = 1, 2, \ldots$, where f_i is the analogous bet on the ith coin. Our subjective ergodic theorem states that the act $f^\star$ is well-defined off a $\succcurlyeq$-null set of states, and that $f_i \sim f^\star$. This roughly says that the decision maker perceives the uncertainty about how the first coin might turn up as 'equivalent' to an uncertainty about the limiting frequency of successive coin tosses.

To motivate the proof of this theorem (in the special case of exchangeability), think of the set of finite permutations Π as a group of transformations on the state space Ω. Exchangeability is precisely the assumption that indifference is preserved under group action: for any event E and permutation π, the decision maker is indifferent between betting on E and betting on the event $\pi(E)$ consisting of all π-permutations of elements of E. This, plus our other basic conditions, is sufficient to establish the conclusion of the theorem, namely that any act is indifferent to its frequentist limit.

As its name suggests, this theorem bears close analogy with the standard ergodic theorem. However, the arguments used to prove that theorem are inapplicable in our context since they fundamentally rely on the existence of a probability measure on the state space. Our starting point, by contrast, is a preference that may be incomplete and/or fail probabilistic sophistication. In fact, our setting is completely deterministic; probabilities emerge as a consequence exchangeability and some basic properties of the preference.

Theorem 2 provides a parametric representation of preferences. We show that there is a partition $\{E^\theta\}_{\theta\in\Theta}$ of the state space with the following properties. Each E^θ admits a unique i.i.d. distribution P^θ such that, off a null event, $f^\star(\omega) = \int_\Omega f\, dP^{\theta(\omega)}$, where $\theta(\omega)$ is the component of the partition to which ω belongs (*i.e.,* the unique θ such that $\omega \in E^\theta$). That is, the expected utility of an act conditional on parameters is precisely the limiting frequency of its payoff at a given state as this state is transformed by permutations. In fact, $f^\star$ is nothing but the certainty-equivalent value of f given θ, and hence constant on each E^θ.

Our third main result, Theorem 3, obtains a new frequentist characterization of subjective expected utility. We show that a decision maker with *ergodic* preference must be a subjective expected utility maximizer with i.i.d. beliefs whose marginal on any single experiment coincides with the empirical frequencies. Ergodicity is a learnability condition, namely that the decision maker believes that, off a $\succcurlyeq$-null set, observing the state ω conveys nothing useful in inferring the value of the parameter. Note that the crucial axioms in Savage's framework, such as completeness or the sure-thing-principle, are not assumed. Rather, our theorem shows that a learnability condition implies these normative properties.

Our final main result, Theorem 4, is a *subjective sufficient statistic theorem*. Given transitivity, we have $f \succcurlyeq g$ if and only if $f^\star \succcurlyeq g^\star$. Since each state ω "belongs to" a unique parameter $\theta(\omega)$, we may define a parameter-based act directly on states as $F\big(\theta(\omega)\big) \equiv \int_\Omega f\, dP^{\theta(\omega)}$. Combining these observations with the last theorem, we have:

$$f \succcurlyeq g \iff F\big(\theta(\cdot)\big) \succcurlyeq G\big(\theta(\cdot)\big).$$

This says that the i.i.d. parametrization Θ is a sufficient statistic for the class of exchangeable preferences, in the sense that in comparing any two acts f and g, it is enough to compare the corresponding parameter contingent acts F and G.

Collectively, our results show that, under exchangeability, one can narrow the difference between various decision making paradigms to how they deal with uncertainty about a common set of parameters. The classic de Finetti's theorem is the special case where the decision maker deals with this uncertainty by introducing a prior on parameters and using the expected utility criterion. But our framework covers more. First, Theorem 5 characterizes the set of events in S treated as unambiguous by all decision makers, irrespective of their attitude towards ambiguity, as those whose empirical frequencies do not depend on the state, off a $\succcurlyeq$-null set. We then discuss consider specific examples of decision criteria, including Bewley (1986)'s model of incomplete preferences and Gilboa and Schmeidler (1989)'s MEU criterion. Decision procedures in classical statistics and econometrics may be similarly incorporated. These models, while explicitly non-Bayesian, typically satisfy our weak auxiliary assumptions and are thus covered by our theorems.

⋆ ⋆ ⋆

Our paper provides, among other things, a link between subjective preferences and objective, frequentist probabilities. To put this contribution in context, it is useful to outline the fundamental tenets underlying de Finetti's conception of subjective probability:[4] (1) *Subjectivity:* probabilities are a decision maker's mental construct, revealed in his observed choices, to help him make sense of his environment; (2) *Exchangeability:* an organizing principle embodying the decision maker's subjective similarity judgement, connecting subjective beliefs with objective frequencies; (3) *Bayesianism:* the decision maker has complete, coherent ranking of all acts.

Our model fully incorporates subjectivity and exchangeability. In particular, probabilities have no objective value, but exist only as cognitive constructs in the mind of the decision maker. The third tenet of de Finetti's methodology, Bayesian beliefs, is both different in nature and, arguably, more questionable. Our results indicate that exchangeability ensures the existence of parameters P^{θ}, and little else. De Finetti's representation requires, in addition, that the decision maker has a prior on the set of parameters, a prior about which exchangeability and frequencies have nothing to say. Thus, the question whether a decision maker has such a prior is logically and conceptually distinct from exchangeability and its core justification as a bridge between subjective beliefs and objective information.

Our approach makes it possible to provide a new perspective on concepts like "objective probabilities" and "true parameters" that are in common use in statistics, economics, and decision theory. An intuitive idea is to define objective probability in terms of frequencies. For example, the probability of *Heads* is the frequency with which *Heads* appear in a sequence of coin tosses. This intuition quickly founders upon the observation that there are

[4]These appear, in one form or another, in most of his writings. Our exposition draws from de Finetti (1937) and de Finetti (1989, originally published in 1931).

uncountably many sequences where the limiting frequency of *Heads* does not converge.[5]

Given our theorems, any decision maker with exchangeable preference must believe that frequencies are well-defined off a subjectively null event. For such decision maker, the definition of objective probabilities as frequencies is meaningful. Needless to say, exchangeability, being a notion of similarity of experiments is, of course, subjective since different decision makers may hold different views about what is and isn't similar.[6]

2 Subjective Ergodic Theory

This section develops a subjective reformulation of a fundamental result in probability theory, *the ergodic theorem*. Roughly, this classic theorem studies the extent to which global properties of a dynamical system can be inferred by tracing the evolution of a single state within that system. We will show that ergodic theory, properly reformulated, can be a powerful tool in the study of exchangeability, de Finetti's theorem, and the foundations of subjective probability. Although its implications for decision theory are our main focus, subjective ergodic theory may be of broader interest. For this reason we introduce it here in abstract terms, postponing connections to decision theory to the next section.

⋆ ⋆ ⋆

Our primitive is a binary relation $\succcurlyeq$ on acts defined on a state space Ω. We assume Ω to be Polish, *i.e.,* a complete separable metrizable space with the Borel σ-algebra Σ. An *act* is any bounded measurable function:

$$f : \Omega \to \mathbb{R}.$$

[5]And uncountably many sequences with limiting frequency of *Heads* equal to a, for any predetermined value of a, and so on.

[6] According to de Finetti (1989, p. 174), a proposition is objective if: "I always know in what circumstances I must call such a proposition true, and in what others false. My calling it true or false implies nothing about my state of mind, signifies no judgement, has no conceptual value." A subjective proposition, on the other hand, is one which "no experience can prove [...] right, or wrong; nor, in general, could any conceivable criterion give any objective sense to the distinction [...] between right and wrong."

Let $\mathcal{F}$ denote the set of all acts. As usual, we identify a real number r with the constant act that pays r regardless of the state. The binary relation $\succcurlyeq$ on $\mathcal{F}$, which we interpret as a preference, is assumed to satisfy the following conditions.

Assumption 1 (Preorder) *$\succcurlyeq$ is reflexive and transitive.*

Note that we do not require preferences to be complete, so preferences of the type considered by Bewley (2002) or those implicit in classical statistical procedures are allowed.

Next we introduce the usual monotonicity assumption (Savage's P3):

Assumption 2 (Monotonicity) *If $f(\omega) \geq g(\omega)$ for all $\omega \in \Omega$, then $f \succcurlyeq g$.*

We finally introduce a continuity assumption on preferences. Given an act f and a sequence of acts $\{f^n\}$ in $\mathcal{F}$, we write $f^n \to f$ to mean that the event $\{\omega \in \Omega : \lim_n f^n(\omega) \neq f(\omega)\}$ is $\succcurlyeq$-null.[7]

Assumption 3 (Continuity) *Suppose that for a given pair of acts $f, g \in \mathcal{F}$ there are sequences $\{f^n\}$, $\{g^n\}$ such that: (i) $f^n \to f$ and $g^n \to g$; (ii) $|f^n(\omega)| \leq b(\omega)$ and $|g^n(\omega)| \leq b(\omega)$, for all ω and some $b \in \mathcal{F}$; and (iii) $f^n \succcurlyeq g^n$. Then $f \succcurlyeq g$.*[8,9]

For the remainder of the paper, a *preference* is any binary relation on $\mathcal{F}$, satisfying Assumptions 1-3.

* * *

[7]We use the standard definition of null events: $E \subset \Omega$ is *$\succcurlyeq$-null* if for all acts f, g and h:

$$\begin{bmatrix} f(\omega), & \text{if } \omega \in E \\ h(\omega), & \text{if } \omega \notin E \end{bmatrix} \sim \begin{bmatrix} g(\omega), & \text{if } \omega \in E \\ h(\omega), & \text{if } \omega \notin E \end{bmatrix}.$$

[8]It is worth noting that, in the special case of expected utility preferences, this condition is equivalent to countable additivity. See Lemma A.24 in the Appendix.

[9]Our continuity assumption is similar to Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003)'s B3. They require that, if $f^n \to f$ and $g^n \to g$ pointwise and $f^n \succcurlyeq g^n$ for each n, then $f \succcurlyeq g$. Note that they do not require the sequences to be bounded by a function b.

A *transformation* is a measurable function $\tau : \Omega \to \Omega$. Fix an act f and a state ω. Repeated applications of the transformation τ generates a sequence of states $\omega,\ \tau\omega,\ \tau^2\omega, \ldots$ The *empirical limit of f (under τ)* is the function

$$f^\star(\omega) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right),$$

whose domain is the set of states where this limit exists. The central question is the relationship between the original act f and its frequentist limit $f^\star$.

The central condition is that the preference $\succcurlyeq$ is τ-*invariant*: for all acts $f \in \mathcal{F}$ and integers $n = 1, 2, \ldots$

$$f \sim \frac{1}{n} \sum_{j=0}^{n-1} f \circ \tau^j. \tag{2}$$

We provide extensive motivation and interpretation of this condition in the context of exchangeability in the next section. A related condition is ergodicity: a preference $\succcurlyeq$ is τ-*ergodic* if for every event E such that $\tau(E) = E$, either E or E^c is $\succcurlyeq$-null.

Theorem 1 (The Subjective Ergodic Theorem) *The following conditions are equivalent:*

1. *$\succcurlyeq$ is τ-invariant;*

2. *For every act f, the empirical limit $f^\star$ is well-defined off a $\succcurlyeq$-null event.*

In this case, $f^\star \sim f$,[10] *and $f^\star$ is τ-invariant, that is, $f^\star(\tau\omega) = f^\star(\omega)$, whenever the limit exists.*

If $\succcurlyeq$ is τ-ergodic, then $f^\star$ is constant except in a $\succcurlyeq$-null set.[11]

We conclude this section by comparing Theorem 1 with the classic Birkhoff ergodic theorem.[12] This classic theorem takes as primitive an objective

[10]Extend $f^\star$ arbitrarily at ω's where the limit does not exist.

[11]The proofs of this and all other theorems are in the appendix.

[12]For references on the standard ergodic theorem, see Billingsley (1995) and, for the related idea of ergodic decomposition, Löh (2006).

probability measure P on Ω, and the invariance condition takes the form $P(E) = P(\tau E)$ for every event E. The two theorems obviously coincide if $\succcurlyeq$ is *an expected utility preference* with a unique subjective prior.[13] Theorem 1, on the other hand, takes as primitive a binary relation that is only required to be transitive, monotone and continuous, but may otherwise fail completeness, the sure thing principle, or the axiom of weak comparative probability. We make no appeal to decision criteria—*e.g.,* the various forms of independence or ambiguity aversion—that deliver probabilities in standard models. Nevertheless, the theorem implies that the preference treats as indifferent an act f and its empirical limit $f^\star$. As Theorem 2 below illustrates, this leads to probabilities as a conclusion, rather than part of the primitives of the model.

If we apply our invariance condition (2) to the indicator function of an event E we get:

$$1_E \sim 1_E \circ \tau. \tag{$*$}$$

If $\succcurlyeq$ is an expected utility preference with subjective probability P, this condition becomes $P(E) = P(\tau E)$, which is the invariance requirement in the classic ergodic theorem. In this case, (*) implies our seemingly stronger condition (2). But when $\succcurlyeq$ is not an expected utility preference, then (2) is potentially stronger than (*).

The reader should keep in mind that condition (2) *characterizes* those preferences for which the conclusions of the ergodic theorem hold. Whether or not one accepts this condition, Theorem 1 says that it is central for a subjective theory of similarity of the sort we develop here. At a minimum, the theorem clarifies what it takes to ensure results such as the existence of empirical limits, sufficient statistics, or the derivation of statistical ambiguity from frequencies. In the next section we make a case why a condition like (2) is normatively compelling.

[13] That is, $\succcurlyeq$ rank acts according to: $f \succcurlyeq g \iff \int_\Omega f\, dP \geq \int_\Omega g\, dP$, with respect to some subjective probability P.

3 Exchangeability, Similarity, and Subjective Probability

For the remainder of the paper we assume the state space to have a product structure: $\Omega = S \times S \times \cdots$, where S is a Polish space endowed with the Borel σ-algebra $\mathcal{S}$. Here, an outcome s may represent something as simple as the result of tossing a coin, or as complex as an observation of an elaborate scientific experiment.[14] The state space Ω reflects a sequence of such experiments indexed by "times" $t = 1, 2, \ldots$[15]

Some of our results will refer to the set $\mathcal{F}^{FB} \subset \mathcal{F}$ of *finitely-based acts* which depend on finitely many coordinates.[16] In Section A.1 we show that our results hold for broader set of acts that satisfy a regularity condition.

We will also interpret acts as either real valued and assume risk neutrality, or that they are directly measured in utils. One may justify the latter interpretation by thinking of more primitive acts taking values in an unmodeled space of consequences and von Neumann-Morgenstern utility mapping these consequences to utils.[17]

[14] We use the term 'experiment' loosely without any presumption of active experimentation on the part of the decision maker. Thus, an econometrician passively collecting evidence is observing the results of experiments in our sense.

[15] Formally, we study a static choice problem where the state space has a product structure. Of course, the study of de Finetti-like representations is motivated by learning from experiments that repeat over time. We briefly discuss learning in Section 7.3

[16] Formally, an act f is finitely-based if there is an integer N, depending on f, such that f is measurable with respect to the algebra generated by events of the form $\{\omega : s_n \in A\}$ for $n = 1, \ldots, N$ and (measurable) $A \subset S$.

[17] This interpretation does entail a loss of generality as it imposes joint restrictions on the space of consequences and the utility function. Although it is possible to make explicit sets of conditions to justify this interpretation, these conditions would distract us from the main points of this paper. Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003) show that it is possible to consider a convex combination of acts and values in utils as we assume here in a purely subjective framework.

3.1 Exchangeability

3.1.1 Formal Definition

The central property of preferences which we wish to study is *exchangeability*. Let Π denote the set of finite permutations $\pi : \mathbb{N} \to \mathbb{N}$. Given an act f and a permutation $\pi : \mathbb{N} \to \mathbb{N}$, the act $f \circ \pi$ is defined as: $f \circ \pi(s_1, s_2, ...) = f(s_{\pi(1)}, s_{\pi(2)}, ...)$.

Exchangeability is a formal expression of the decision maker's perception that experiments are similar. Consider the stylized setting of coin tosses, where each s_n is either *Heads* or *Tails*. Let f be the act that pays \$1 if the first coin toss turns up *Heads* and zero otherwise. If π denotes the permutation of the indices 1 and 2, then $f \circ \pi$ is just the act that bets \$1 on the second coin. Exchangeability then says that the decision maker is indifferent between betting on the first and second experiments.

More generally, a *minimal* requirement for exchangeability is that for every act $f \in \mathcal{F}$ and permutation $\pi \in \Pi$,

$$f \sim f \circ \pi. \tag{3}$$

Under expected utility, condition (3) is sufficient to establish the central conclusion of de Finetti's theorem, namely linking subjective probability and frequencies. But de Finetti's notion of exchangeability, being conceived for the expected utility context, loses much of its force in our more general setting. To recover the link between preferences and frequencies, Theorem 1 suggests that a condition in the spirit of (2) is needed. In our case it is natural to require:

Assumption 4 (Exchangeability) *A preference $\succcurlyeq$ is* exchangeable *if for every act $f \in \mathcal{F}$ and permutations $\pi_1, \ldots, \pi_n$*

$$f \sim \frac{f \circ \pi_1 + \cdots + f \circ \pi_n}{n}. \tag{4}$$

This reduces to condition (3) when $n = 1$. Assuming expected utility, exchangeability can be derived from the weaker condition (3) with a simple application of the independence axiom.[18]

[18]The condition characterizing preferences that satisfy our central result, the subjective

3.1.2 Motivation

To provide a normative motivation for condition (4), think of an exchangeability relationship as the decision maker's subjective theory, or model, of similarity. The need for similarity arises because the decision maker recognizes that the outcome of the coin toss is determined by the interaction of a multitude of poorly understood factors that he either cannot or is unwilling to precisely model. This, after all, is why the coin sometimes comes up *Heads*, and some other times *Tails*!

The fact that there are poorly understood factors that can influence outcomes is inherent even in mundane settings like coin tosses. De Finetti's insight is that exchangeability is what enables the decision maker to pool information across experiments through the subjective judgment that these poorly understood factors affect experiments symmetrically.

The minimal requirement $f \sim f \circ \pi$ goes some distance in formalizing our intuition of the decision maker's perception of similarity. But it is wholly inadequate to capture the crucial insight in de Finetti's classic theorem, namely to identify parameters that we can meaningfully link to frequencies and interpret as sufficient statistics for decision making. In fact, our exchangeability (4) requirement is essentially necessary for our results, as shown in Corollary 3. Thus, the interpretations of parameters as sufficient statistics would not hold with weaker notions of exchangeability.

Interpreting exchangeability as the decision maker's subjective model of similarity, it is natural to require that he be committed to whatever model he has chosen. We formalize this commitment by appealing to the minimal version of the independence axiom implicit in (4). Independence is applied only to the exchangeability indifferences, and only in so far as simple averages are concerned. This is a weak requirement in that it is entirely consistent with, for instance, ambiguity about the parameters.

In sum, what we rule out is an incoherent decision maker who views ex-

---

ergodic theorem, is (2). We use (4) because it is stated in terms of permutations which we view as primitive, while (2) is stated in terms of abstract transformations. In one of their theorems, Epstein and Seo (June 2008) use the condition that, for every $f \in \mathcal{F}, \pi \in \Pi$, and $\alpha \in [0,1]$, $f \sim \alpha f + (1-\alpha) f \circ \pi$. Under our weak assumptions on preferences, this condition neither implies nor is implied by (4). See Section 5.3.

periments as similar, yet he is not sure that they are. Such an incoherent perception of what constitutes similar experiments is interesting as a behavioral bias, but difficult to imagine as a basis for statistical and econometric practice, or as a useful ingredient in economic models.

3.1.3 Exchangeability, Frequencies, and Patterns

Before closing this section we restate Theorem 1 in the special context of exchangeability. The transformation in this case is the shift, defined by:

$$T(s_1, s_2, \ldots) = (s_2, s_3, \ldots).$$ [19]

Corollary 1 *For every exchangeable preference* $\succcurlyeq$ *and every* $f \in \mathcal{F}^{FB}$*, the empirical act*

$$f^\star(\omega) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j \omega\right)$$

is well-defined off a $\succcurlyeq$*-null set and* $f \sim f^\star$*.*

If $\succcurlyeq$ *is (T-)ergodic, then* $f^\star$ *is constant outside a* $\succcurlyeq$*-null set.*

In the stylized coin tossing example, where f is a \$1 bet on *Heads* in the first toss, $f(T^j(s_1, s_2, ...))$ is simply the act that pays \$1 if the $j+1$ toss turns up *Heads* and zero otherwise. If $\succcurlyeq$ is exchangeable, then $f^\star$ is well-defined and $f \sim f^\star$. Note that $f^\star(s_1, s_2, \ldots)$ is nothing but the limiting frequency of *Heads* in this sequence. Thus, $f \sim f^\star$ says that the decision maker is indifferent between a bet on *Heads* in a single toss and an act that pays an amount equal to the frequency of *Heads* (recall that acts pay in utils, so risk aversion is not an issue).

At first sight, this may look like 'magic:' how is it possible that a conclusion about an objective entity, namely the empirical limit $f^\star$, can be derived from a purely subjective preference relation? There is, of course, no magic

[19] Note that the shift is not a permutation, but under continuity and our assumption that acts are finitely based, shifts are preference-equivalent to permutations. Thus, in the case of a bet on the first coin, shifting it to the ith coordinate can be achieved by a permutation that takes the first $i-1$ coordinates and 'pushes' them to some finite, but remote index j. This procedure can be replicated for all finitely-based acts.

here. In the coin tossing case, there are uncountably many sequences where the empirical frequency of *Heads* does not exist. Although the value $f^{\star}(\omega)$ for a given ω is entirely objective, it is the *subjective judgement* of the similarity of the experiments that implies that the decision maker views the set of sequences where the limit fails to exist as $\succcurlyeq$-null.

The reader may wonder whether acts defined in terms of limiting frequencies have any practical relevance in a decision making context. Our result that $f \sim f^{\star}$ says that the decision maker decomposes the uncertainty he faces into a systematic 'pattern,' and idiosyncratic 'noise' around that pattern. Whether a coin turns up *Heads* or *Tails* in a particular toss is idiosyncratic, being the outcome of a multitude of complex, poorly understood factors. A decision maker with exchangeable preference views these outcome as fluctuations around a systematic pattern which he subjectively identifies with the limiting frequency that would have obtained if the experiment were repeated under similar conditions. That this pattern is mathematically represented as an infinite limit $f^{\star}$ is, in a sense, incidental. This mathematical form just happens to be what we need to represent the intuitive link between patterns and frequencies. We can alternatively characterizes these patterns as subjective probabilities, as we do next.

3.2 Frequencies as Subjective Probabilities

Our next theorem provides a parametrization of exchangeable preferences. First we introduce some standard definitions: given an event $A \subset \Omega$ and permutation $\pi \in \Pi$, define πA to be the event consisting of all states $(s_{\pi(1)}, s_{\pi(2)}, ...)$ for some $(s_1, s_2, ...) \in A$. An event A is exchangeable if $A = \pi A$ for every permutation. A probability distribution P over Ω is exchangeable if $P(A) = P(\pi A)$ for every event A and permutation π. For any $\theta \in \Delta(S)$, define θ^{∞} to be the product distribution of θ over Ω. We say that P is i.i.d. if $P = \theta^{\infty}$ for some θ.

Next we introduce an abstract notion of parametrization:

Definition 1 *A* parametrization*, with index set* Λ*, is a set of pairs* $\left\{\left(E^{\lambda}, P^{\lambda}\right)\right\}_{\lambda \in \Lambda}$ *where the* E^{λ}*'s form a partition of* Ω *and* P^{λ} *is a probability measure with* $P^{\lambda}(E^{\lambda}) = 1$.

A parametrization is exchangeable (i.i.d.) *if each* P^λ *is.*

We will typically refer to a parametrization by its index set Λ. Also, define $\lambda(\omega)$ to be the value of λ for which $\omega \in E^\lambda$.

Theorem 2 (Subjective Parametrization) *There is an i.i.d. parametrization* $\left\{\left(E^\theta, P^\theta\right)\right\}_{\theta\in\Theta}$ *with the following property: For every exchangeable preference* $\succcurlyeq$ *and every* $f \in \mathcal{F}^{FB}$*, for* P^θ*-almost every* ω

$$f^\star(\omega) = \int_\Omega f\, dP^{\theta(\omega)}.$$

Obviously, the i.i.d. measures P^θ are unique; non-uniqueness stems only from the fact that the E^θ's are not unique.

The theorem says that we can think of the values of the empirical act $f^\star$ equivalently as certainty equivalents with respect to uniquely identified subjective probabilities. Note that the probabilities P^θ's emerge as consequence of exchangeability in an otherwise deterministic setting. Note also that, under expected utility, Hewitt and Savage (1955)'s theorems imply that there is an i.i.d. parametrization, but are silent about the existence of frequentist acts $f^\star$ and their relationship to the expected value of f under P^θ. Without expected utility, the Hewitt and Savage's theorems have no force.

3.3 Subjective Probabilities as Frequencies

Subjective probability in the classic Savage (1954) framework is derived from axioms on preferences that are justified based on their normative appeal. The most critical of these axioms are completeness, the sure thing principle, and the axiom of comparative probability.[20]

Using our framework, we are able to derive subjective probability based on learning foundations: Any decision maker with an *ergodic* preference must be an expected utility maximizer, with subjective beliefs given by the empirical frequencies.[21] Ergodicity roughly means that the decision maker believes he

[20]These are Savage's P1, P2 and P4, respectively.

[21]To our knowledge, the only other attempt to relate subjective probabilities to frequencies appears in Hu (2008), who works in a von Neumann-Morgenstern setting and uses arguments quite distinct from ours.

will not learn anything new by observing additional data. Thus, a learning assumption is shown to imply the above mentioned Savage axioms.

First we need some formal definitions:

- $E \subset \Omega$ is $\succcurlyeq$-*trivial* if either E or E^c is $\succcurlyeq$-null;
- E is *invariant* if $T(E) = E$;
- $\succcurlyeq$ is *ergodic* if it is exchangeable and all invariant sets are $\succcurlyeq$-trivial;

To appreciate the definition, consider a repeated coin toss and suppose that a decision maker believes there are only two possible biases of the coin: θ_1 and θ_2. Let E^{θ_1} and E^{θ_2} denote the sets of all sequences with limiting frequencies θ_1 and θ_2, respectively. Then E^{θ_1} and E^{θ_2} are both invariant, but not $\succcurlyeq$-trivial. Here the lack of ergodicity corresponds to the existence of events the uncertainty about which can be resolved via knowledge of the long run frequencies. Given this, there is little we can say about how a decision maker may evaluate the act f_1 that pays \$1 if the first coin turns *Heads* and zero otherwise. The decision maker may have a Bayesian prior on θ_1 and θ_2, ambiguous beliefs, or an incomplete preference. An ergodic preference, on the other hand, is one for which such knowledge is of no value. Roughly, a decision maker with an ergodic preference believes he has learned all that can be learned about the uncertainty he faces.

For a subset $A \subset S$ and state ω, the number $1^\star_A(\omega)$ is simply the frequency of A in ω.[22]

Of course, for a given A and ω, the frequency may either fail to exist or may depend on ω. The *empirical distribution* at ω is the set-function which assigns to each A the value

$$\nu(A, \omega) \equiv 1^\star_A(\omega),$$

if this value exists, and is not defined otherwise. For the next theorem, we shall require the following strengthening of monotonicity:

[22] Here, we simplify notation by speaking of an event $A \subset S$ to refer to $A \times S \times \cdots \in \Sigma$.

Assumption 5 (Strict Monotonicity for Consequences) *For every $x, y \in \mathbb{R}$,*

$$x > y \implies x \succ y.$$

Theorem 3 *If $\succcurlyeq$ is ergodic, then:*

- *There is an event Ω' with $\succcurlyeq$-null complement, such that the empirical distribution $\nu(\cdot) = \nu(\cdot, \omega)$ is a well-defined probability distribution on $\mathcal{S}$ that is constant in $\omega \in \Omega'$; and*

- *$\succcurlyeq$ is an expected utility preference with subjective probability P that is i.i.d. with marginal ν.*

A key difficulty in proving this theorem is ensuring that, unlike in Theorem 1, the set Ω' does not depend on the act f.

This theorem is not a generalization of the Savage framework, but a distinct derivation of subjective probability based on different principles. Let $\succcurlyeq$ be a preference that satisfies our basic assumptions:

- In the Savage approach, one first adds other normatively motivated axioms on $\succcurlyeq$, such as completeness and the sure thing principle, to derive a subjective expected utility representation. If, in addition, we assume that $\succcurlyeq$ is exchangeable, then the subjective probability thus derived is necessarily exchangeable, and one may use standard tools from probability theory (*e.g.,* the law of large numbers) to derive results about the properties of the empirical frequencies.

- Our theorem takes a completely different route: we steer clear from any additional normative axiomatic assumption, and focus instead on what the decision maker believes he can learn from observations (ergodicity). We then use the subjective ergodic theorem and other results to derive the empirical measure and show that the decision maker must have a subjective expected utility preference with subjective belief equal to this measure. This in turn implies that the preference must satisfy the more controversial of Savage's axioms, such as completeness and the sure thing principle.[23]

[23]The theorem can be applied under a weakening of the assumption of ergodicity. Call a

4 Sufficient Statistics, Inter-subjective Agreement, and de Finetti's Theorem

4.1 The Subjective Sufficient Statistic Theorem

What makes a parametrization useful is that it captures all that is relevant for the decision maker's preference. This is closely related to the concept of sufficiency in mathematical statistics. Recall that a measurable function $\kappa : \Omega \to A$, where A is an abstract measurable space, is a *sufficient statistic* for a family of probability distributions $\mathcal{P}$ if the conditional distributions $P(\cdot \,|\, \kappa)$ do not depend on P.[24] Roughly, κ is sufficient if it captures all the relevant information contained in a state ω: given knowledge that $\kappa = \bar{\kappa}$, no further information about ω is useful in drawing an inference about P.

In generalizing this intuition to preferences in a subjective framework, two difficulties arise: First, since distributions in the statistics literature are objective, sufficiency notions expressed in terms of objective probabilities need not capture what is relevant for preferences. Second, sufficiency in the statistics literature is inherently tied to probabilities, making it inapplicable to preferences that may fail to be complete or probabilistically sophisticated.

To make the idea of sufficiency for preferences formal, given a parametrization Λ, define $\mathcal{F}_\Lambda \subset \mathcal{F}$ to be the set of acts measurable with respect to $\{E^\lambda\}_{\lambda \in \Lambda}$.[25] Define the mapping

$$\Phi_\Lambda : \mathcal{F}^{FB} \to \mathcal{F}_\Lambda \qquad by \qquad \Phi_\Lambda(f)(\omega) = \int_\Omega f \, dP^{\lambda(\omega)}. \tag{5}$$

We extend the notion of sufficiency to a subjective setting by defining:

preference Σ'-*ergodic* if it is ergodic with respect (Ω, Σ'), where Σ' is a sub-σ-algebra of Σ. That is, we require $\succcurlyeq$ to be exchangeable and all invariant sets in Σ' are $\succcurlyeq$-trivial. Then, with little change in the proof, we can conclude that $\succcurlyeq$ is an expected utility preference over acts that are Σ'-measurable.

[24] We are abstracting from measure theoretic issues regarding the definition of conditional probabilities for expository simplicity. See, for instance, Billingsley (1995).

[25] Formally, the sub-σ-algebra of Σ generated by the E^λ's.

Definition 2 *A parametrization Λ is* sufficient *for a set of preference relations $\mathcal{E}$ if for each $\succcurlyeq \in \mathcal{E}$ and finitely-based acts f and g*

$$f \succcurlyeq g \iff \Phi_\Lambda(f) \succcurlyeq \Phi_\Lambda(g). \tag{6}$$

Thus, a parameterization is sufficient if the restriction of a preference to the subset of acts $\mathcal{F}_\Lambda$ is sufficient to determine the entire preference, simultaneously for all preferences in the class $\mathcal{E}$.

Two things should be noted about the definition. First, the use of the expected utility criterion implicit in the definition of Φ_Λ *defines* what we mean by a "parameter:" once the parameter is known, any remaining uncertainty must be treated as risk. Without this criterion, sufficiency loses its meaning as a device for compressing information. Second, subjective sufficiency, like its counterpart in statistics, is a property of a class of preferences. It makes little sense to talk about sufficiency for a single preference. For example, if $\succcurlyeq$ is an expected utility preference with subjective belief P, then the trivial parametrization (Ω, P) is "sufficient" for $\succcurlyeq$, trivially.

To establish the next theorem, we need an additional assumption of mainly technical nature:

Assumption 6 *For every $\succcurlyeq$-non-null event E there is an i.i.d. distribution P such that $P(E) > 0$.*

If we think of a preference $\succcurlyeq$ as an aggregator of parameter-contingent preferences, then the assumption says that $\succcurlyeq$ cannot be so badly behaved as to treat as non-null an event that is conditionally null at each parameter.

Theorem 4 (Subjective Sufficient Statistic) *The parametrization Θ in Theorem 2 is sufficient for the set of exchangeable preferences satisfying Assumption 6.*

4.2 Parameter-based Acts and de Finetti's Theorem

Given an act $f \in \mathcal{F}$, its *parameter-based reduction* is the act:

$$F : \Theta \to \mathbb{R} \quad \text{where} \quad F(\theta) = \int_\Omega f \, dP^\theta.$$

Conversely, for any act $F : \Theta \to \mathbb{R}$ there corresponds an equivalence class of state-based acts whose reduction is F. Here, F expresses the state-based act f in terms of the parameters, and is essentially the same as $\Phi_\Theta(f)$ except that we write it directly in terms of the parameters θ, rather than the primitive states ω.

We are now in a position to state de Finetti's classic result. By a *belief on a parametrization* Λ we mean a probability distribution on Ω endowed with the sub-σ-algebra of Σ generated by $\{E^\lambda\}_{\lambda \in \Lambda}$.

de Finetti's Theorem: *An expected utility preference $\succcurlyeq$ is exchangeable if and only if there is a belief μ on Θ such that:*

$$f \succcurlyeq g \iff \int_\Theta F\, d\mu \geq \int_\Theta G\, d\mu.$$

De Finetti's theorem says that:

1. Θ is *sufficient* for the class of exchangeable expected utility preferences;
2. the decision maker resolves uncertainty about the parameters using the expected utility criterion.

Our theorems show a stronger version of the first statement and entirely drop the second: on the one hand, Θ is sufficient for all exchangeable preferences, whether or not they have expected utility representation. On the other hand, we allow preferences where decision makers need not reduce uncertainty about parameters to risk.

We conclude by noting that the relationship between de Finetti's theorem and sufficient statistics has been discussed in the literature in the context of expected utility preferences. See, for instance, Diaconis and Freedman (1984), Lauritzen (1984), and Diaconis (1992). The general formulation we have here is new.

4.3 Objective Probabilities as an Inter-subjective Consensus

Our results provide a new perpective on the question: *To what extent would subjectivist decision makers' beliefs (dis-) agree?* In particular, are there aspects of beliefs that all subjectivists agree on? In general, little can be said in this regard. However, under exchangeability and Assumption 6, we can show that preferences cannot contradict a Bayesian consensus.

Let Υ is the set of all exchangeable probability distributions. In what follows, by an *exchangeable Bayesian* we shall mean an expected utility decision maker with exchangeable subjective belief P.

Corollary 2 (Inter-subjective Consensus) *If $\succcurlyeq$ is any exchangeable preference satisfying Assumption 6, then for all acts f, g,*

$$\int_\Omega f\,dP \geq \int_\Omega g\,dP,\ \forall P \in \Upsilon \implies f \succcurlyeq g.$$

In words, a decision maker with an exchangeable preference agrees with the Bayesians' consensus ranking of acts, whenever such consensus exists. The corollary amounts to asserting that for every exchangeable preference $\succcurlyeq$, the implied preference on parameter-based acts is monotone.[26]

The corollary may be interpreted as saying that the P^θ's are "objective" among all decision makers who regard the experiments as similar, in the sense that they all share a common assessment of the probabilities conditional on knowing the parameters. Subjectivity enters only in the way Bayesians resolve uncertainty about the parameters. For a Bayesian, his beliefs conditional on parameters are pinned down by exchangeability and frequencies. On the other hand, how he forms beliefs about the relative weights of parameters is determined, as in all Savage-style models of decision making, by factors that originate outside the model.

[26]For a proof, suppose that all exchangeable Bayesians prefer f to g. This, in particular, implies that $\int f\,dP^\theta \geq \int g\,dP^\theta$ for each $\theta \in \Theta$. From this it follows, that for each ω outside the complement of a null set $f^\star(\omega) = E_{P^\theta} f \geq E_{P^\theta} g = g^\star(\omega)$. By monotonicity, we have $f^\star \succcurlyeq g^\star$, and by Theorem 1 and transitivity, $f \succcurlyeq g$.

5 Exchangeability and Ambiguity

5.1 Statistical Ambiguity

According to Theorem 3, a decision maker with an ergodic preference views all events as unambiguous. Our next theorem identifies a subset of events the decision maker considers to be, in a sense we make precise, *statistically unambiguous* even though the preference may be incomplete and/or violate probabilistic sophistication. For our next result we assume a finite outcome space S and focus on the set of acts $\mathcal{F}_1$ that depend only on the first coordinate.[27]

Definition 3 *An event $A \subset S$ is* $\succcurlyeq$-statistically unambiguous *if there is $\Omega' \in \Sigma$ with $\succcurlyeq$-null complement, such that the frequency of A, $\nu(A) = \nu(A, \omega)$, is constant in $\omega \in \Omega'$.*

The crucial part of the definition is the requirement that $\nu(A, \omega)$ is independent of ω off a $\succcurlyeq$-null set. Intuitively, an event A is statistically unambiguous if the decision maker is confident about his assessment of its probability. We interpret this to mean that he is convinced that the empirical frequency of that event along any infinite sample will confirm his subjective belief about the likelihood of that event. Theorem 5 below will validate this interpretation.

Definition 4 *Given a family of subsets $\mathcal{C} \subset 2^S$, a partial probability ν on $\mathcal{C}$ is a function $\nu : \mathcal{C} \to [0, 1]$ such that there is a probability distribution on S that agrees with ν on $\mathcal{C}$.*

For a partial probability ν and $\mathcal{C}$-measurable function $f \in \mathcal{F}_1$, we define the integral $\int f \, d\nu$ in the obvious way.

The theorem characterizes those events that can be declared unambiguous from the learning standpoint. The theorem says nothing about the decision maker's attitude in dealing with the remaining statistical ambiguity.

[27]The restriction to finite S is for technical reasons, while the restriction to $\mathcal{F}_1$ is mainly for expositional reasons.

Theorem 5 (Statistical Ambiguity) *Assume S is finite. For any exchangeable preference $\succcurlyeq$*

- *The set of $\succcurlyeq$-statistically unambiguous events $\mathcal{C} \subset 2^S$ is a λ-system, i.e., a family of sets closed under complements and disjoint unions;*
- *The empirical measure $\nu(\cdot)$ is a partial probability on $\mathcal{C}$; and*
- *For every $\mathcal{C}$-measurable acts $f, g \in \mathcal{F}_1$:*

$$f \succcurlyeq g \iff \int f \, d\nu \geq \int g \, d\nu.$$

Theorem 5 identifies the set of statistically unambiguous events in terms of the partial probability ν. The decision maker has expected utility preference over acts $f \in \mathcal{F}_1$ that are measurable with respect to set of events $\mathcal{C}$ on which ν is defined.

While exchangeability and frequencies pin down the probabilities of events in $\mathcal{C}$, the decision maker's assessment of the probabilities of the remaining events will depend on aspects of his preference beyond the minimal assumptions we impose. Additional aspects of the preference, such as the sure thing principle, completeness, or ambiguity aversion/neutrality become key in determining how the decision maker deals with statistical ambiguity. Our approach here is to determine what can be said on statistical, frequentist, grounds.

Note that $\mathcal{C}$ is only a λ-system. This is consistent with the intuitions that appeared first in Zhang (1999) and Nehring (1999). In fact, Zhang's example, adapted to our setting, would illustrate that $\mathcal{C}$ need not be an algebra. The intuition in terms of empirical frequencies is quite clear: given any state ω, if the empirical frequency of an event A exists, then so would the frequency of its complement. A similar conclusion holds for any two *disjoint* events A and B. But this is all we can conclude in general. See the Appendix for further discussion.

5.2 Ambiguity: *Incomplete Preferences*

The primitive in Bewley (1986)'s model is an incomplete preference $\succcurlyeq$, which he characterizes in terms of a (compact, convex) set of probability measures

C such that:

$$f \succcurlyeq g \iff \int_\Omega f\, dP \geq \int_\Omega g\, dP, \;\; \forall P \in C. \tag{7}$$

Under this criterion, f is preferred to g iff f yields at least as high an expected payoff as g with respect to each and every distribution in C. The set C is interpreted as a representation of the decision maker's ignorance of the "true" probability law governing the observables. Bewley compares this unanimity criterion to classical statistics where an inference is valid if it holds for all distributions in a given class. The set C has been interpreted as representing what is objectively known to the decision maker.[28]

One may think of Bewley's model as one where the set of "parameters" is the set of all probability distributions $\Delta(\Omega)$. As a result, the subjective set of measures C has little structure beyond compactness and convexity. Without exchangeability, or some other structure, Bewley's model permits severe and, arguably, unreasonable forms of incompleteness. For instance, C may include the set $c_{\frac{1}{2}}$ of all Dirac measures on sequences of coin tosses with frequency of heads converging to 0.5. In this case a Bewley decision maker would prefer f over g only if f gives at least as good a payoff as g on each and every such sequence.

Exchangeability imposes a natural structure on the set of parameters and, consequently, on preferences. Rather than allowing all distributions, an exchangeable Bewley decision maker is willing to treat "within-parameter" uncertainty as pure risk. Any remaining ambiguity is due to lack of knowledge of the values of the parameters, expressed by the requirement that C be a subset of $\{P^\theta\}_{\theta\in\Theta}$. An exchangeable Bewley decision maker will therefore have a preference over parameter-based acts give by:

$$f \succcurlyeq g \iff F(\theta) \geq G(\theta), \;\; \forall \theta \in \Theta.$$[29]

In the case where the set of priors is the set $c_{\frac{1}{2}}$ above, an exchangeable Bewley decision maker will have an ergodic preference and, by Theorem 3,

[28] Gilboa, Maccheroni, Marinacci, and Schmeidler (2008) make this interpretation explicit.

[29] Strictly speaking, we should consider the convex hull of C. However, given linearity in probabilities, focusing on the extreme points suffices and has the virtue of simplifying the exposition.

he must have a complete expected utility preference with subjective belief $P^{\theta=1/2}$. We finally note that, under exchangeability, disagreements among Bewley's decision makers center around the values of parameters, in which case his model lends itself more naturally to a classical statistics interpretation.

5.3 Ambiguity: *Malevolent Nature*

Our framework can also accommodate models of ambiguity aversion, such as the variational preference model of Maccheroni, Marinacci, and Rustichini (2006). The variational model includes many well-known ambiguity averse preferences as special cases. We illustrate our point with one important special case, the Gilboa and Schmeidler (1989)'s MEU criterion:

$$f \succcurlyeq g \iff \min_{P \in C} \int_\Omega f \, dP \geq \min_{P \in C} \int_\Omega g \, dP \tag{8}$$

for some compact, convex set of probability measures C. This model and its variants can be interpreted as "games against Nature," where a malevolent Nature changes the true distribution as a function of the choices made by the decision maker.[30]

As with Bewley's model, without the assumption of exchangeability, we may think of the set of "parameters" as all of $\Delta(\Omega)$. As a result, the model can display an unreasonable degree of ambiguity aversion. For example, with the set $c_{\frac{1}{2}}$ of Dirac measure on sequences converging to 0.5, the decision maker is worried about each and every state in this set. Exchangeability, again, introduces a natural structure, with C required to be the closed convex hull of a subset of $\{P^\theta\}_{\theta \in \Theta}$. In this case, the MEU criterion becomes:

$$f \succcurlyeq g \iff \min_{\theta \in C} F(\theta) \geq \min_{\theta \in C} G(\theta)$$

Exchangeability does not rule out ambiguity aversion. For instance, in the case of coin tosses, the decision maker may believe that the true probability of heads is either $\theta = 0.4$ or $\theta' = 0.6$, with each assigned at least probability $\frac{1}{4}$. The set C in this case consists of all distributions on the two-point set

[30]For a critical assessment of the ambiguity aversion literature, see **?**.

$\{\theta, \theta'\}$ assigning probability at least $\frac{1}{4}$ to each element. This is a version of Ellsberg's two-color problem, except that ball colors are now replaced by values of the parameter. Although ambiguity aversion can arise in our model, exchangeability limits it to the value of the parameter; once the parameter is revealed, the decision maker has a standard expected utility preference. Effectively, exchangeability limits what a malevolent Nature can do to harm a decision maker who believes he is facing repeated outcomes of a stochastically invariant phenomenon.

A related paper by Epstein and Seo (June 2008) exclusively studies MEU preferences, due to Gilboa and Schmeidler (1989). As noted earlier, these are a special case of variational preferences which our model covers. They offer two submodels. In their more substantial submodel, Epstein and Seo maintain (3) yet allow, for instance, $\frac{1}{2} f \circ \pi_1 + \frac{1}{2} f \circ \pi_2 \succ f$ for some act f and permutations π_1, π_2, even though $f \sim f \circ \pi_1 \sim f \circ \pi_2$. They interpret the strict preference as the decision maker's perception of a lack of "evidence of symmetry," namely that there are poorly understood factors that make the "true" process non-exchangeable. As we noted in Section 3.1, poorly understood factors are present even in the stylized case of coin tosses, and even assuming expected utility preferences. It is precisely such factors that, after all, cause the coin to turn up differently in different experiments! The only substantive question, then, is how a decision maker incorporates these factors, not whether they exist.

In this submodel, parameters are *sets* of measures, and their framework allows for a very broad range of such "parameters." For example, the set of all Dirac measures on Ω is a possible parameter. Another parameter is the set of all independent distributions, the set of all independent distributions with marginals belonging to $\{.2, .6\}$, to $\{.43, .91\}$, ... etc.[31] In their representation

[31] More precisely, in the case of coin tosses, interpret a closed subset $L \subset [0, 1]$ as a set of coin biases in a single experiment. Given L, define L^∞ to be the set of all product measures $l_1 \otimes l_2 \otimes \cdots$ on Ω determined by tossing the coin $l_n \in L$ at the nth experiment. (Strictly speaking, parameters are the closed convex hull of sets of the form L^∞, but under the MEU criterion only the extreme points matter, and these are just L^∞.) For Epstein and Seo, any *set of measures* taking the form L^∞ for some L is a possible parameter.

Note that the set L^∞ is symmetric, in the sense that, for every permutation π, $l_{\pi(1)} \otimes l_{\pi(2)} \otimes \cdots$ belongs to L^∞ if $l_1 \otimes l_2 \otimes \cdots$ does. However, unless L is a singleton, L^∞ will always contain non-exchangeable distributions. For example, if we take $L = \{0, 1\}$, the decision maker has a probability measure on such "parameters," each of which is, in turn, a set of measures. What determines the support of the decision maker's belief over parameters is, partly, his distaste for ambiguity and, partly, his fear that Nature might have rigged the experiment to be non-symmetric. To sum up, a useful insight of Epstein and Seo's main submodel is that it provides a clear sense of the anomalies that might arise when the exchangeability condition (4) is weakened. In our view, a parametrization where individual parameters are sets, including the set of all sample paths, seems so far removed from the intuitive idea of parameters as useful devices to summarize information. It is difficult to imagine how statistical inference can proceed on this basis.

Their other submodel adds to the Gilboa and Schmeidler axioms the property that, for every $f \in \mathcal{F}, \pi \in \Pi,$ and $\alpha \in [0,1]$,

$$f \sim \alpha f + (1-\alpha) f \circ \pi. \tag{9}$$

This condition neither implies nor is implied by our central exchangeability condition (4).[32] We see no substantive reason to reject our exchangeability condition (4) if one is willing to accept (9). On the other hand, (4) enables us to consider the implications of exchangeability for much broader settings than the MEU special case. Our results are also quite different: Epstein and Seo use the standard Gilboa-Schmeidler axioms and representation to derive the intuitive result that the decision maker's set of priors consists of exchangeable distributions. By contrast, our approach is to avoid

then any sequence of *Heads* and *Tails* is a degenerate distribution that belongs to the "parameter" $L^\infty \equiv \Omega$. Aside from the two sequences where the outcome is constant *Heads* or constant *Tails*, no measure in this parameter is exchangeable. A decision maker who entertains Ω as a parameter is so paranoid that he hedges against each and every sequence of outcomes, voiding the very motivation for using parameters in decision making.

[32]It may be easier to illustrate this by showing that, under our weak assumptions, (9) does not imply: $f \sim \sum_{i=1}^{k} a_i\, f \circ \pi_i$, for permutations $\pi_1, \ldots, \pi_k$ and real numbers $a_1, \ldots, a_k$ in [0,1] with $\sum_{i=1}^{k} a_i = 1$. Our condition (4) is the special case with $a_i = \frac{1}{k}$. To see this, consider the naive argument that derives this condition by applying (9) to $a_k\, f \circ \pi_k + (1-a_k) \sum_{i=1}^{k-1} \frac{a_i}{1-a_k}\, f \circ \pi_i$ and conclude that $f \sim a_k\, f \circ \pi_k + (1-a_k) \sum_{i=1}^{k-1} \frac{a_i}{1-a_k}\, f \circ \pi_i$. The problem is that (9) applies only to acts of the form $f \circ \pi$ for some $\pi \in \Pi$, and $\sum_{i=1}^{k-1} \frac{a_i}{1-a_k}\, f \circ \pi_i$ need not be of this form. Although potentially weaker, (9) may imply (4) under specific functional forms, such as expected utility.

introducing substantive axioms reflecting decision makers' attitude towards ambiguity (*e.g.,* axioms that lead them to be Bayesians, use the MEU or Bewley criterion, or some other method) and focus instead on the implications of exchangeability. Thus, our Theorem 5 identifies the set of events which all exchangeable decision makers view as unambiguous, irrespective of their attitude to ambiguity.

6 Partial Exchangeability

While the concept of exchangeability looms large in de Finetti's conception of subjective probability, it has played a relatively minor role in modern axiomatic decision theory. One reason may be the seemingly stylized, unrealistic behavior it depicts, namely that the parameters are i.i.d. and the preference is invariant with respect to all acts and permutations. For instance, a decision maker with an exchangeable preference is required to view each and every event $A \subset S$ as equally likely in all experiment.

Here we show that our framework can effortlessly accommodate important forms of *partial exchangeability*, thus relaxing the unyielding requirement that the decision maker must view *all* aspects of the experiments as similar.[33] The ensuing discussion also illustrate the value of the general form of the subjective ergodic theorem we present in Theorem 1.

First, our framework can readily accommodate Markov processes, along the lines pursued by Diaconis and Freedman (1980) in a Bayesian setting. In this case parameters are transition kernels rather than i.i.d. distributions.

Second, consider a version of partial exchangeability that requires invariance with respect to only a subset of acts. We model this formally by assuming that exchangeability applies only to a set of acts $\tilde{\mathcal{F}}$ measurable with respect to some sub-σ-algebra $\tilde{\mathcal{S}} \subset \mathcal{S}$. In the case of coin tosses, the 'true' outcome space S may reflect a multitude of factors and their complex interactions, such as the precise distance the coin travels, the force applied, the room temperature, the nature of the surface it lands on, the dynamics of air turbulence it creates, ... etc. Let $\tilde{\mathcal{S}}$ be the algebra of events generated by

[33]To be sure, this was already anticipated by de Finetti (1938), and further developed, in a Bayesian setting, by several authors. See, for instance, Diaconis and Freedman (1984).

the event $H \subset S$ which consists of all those circumstances in which the coin turns up *Heads*, and its complement $T \equiv S - H$. If we require exchangeability on the coarse outcome space $(S, \tilde{\mathcal{S}})$, then all of our results apply without modification. Requiring exchangeability with respect to $\tilde{\mathcal{S}}$, but not $\mathcal{S}$, reflects a decision maker who believes that, although no specific factor need to arise exchangeable in the various experiments, the combination of all factors that collectively cause the coin to turns up *Heads* does.

This coarsening of the set of events corresponds to a "small world" in the sense of Savage (1954). Savage's motivation was that the full set of events $\tilde{\mathcal{S}}$ may be too complex or detailed for the decision maker to usefully specify a preference on, so he opts for a small world $\mathcal{S}$ instead. Our (marginal) contribution is to propose exchangeability as a principle that motivates and guides the choice of a small world. The decision maker may subjectively view the large world $\tilde{S}$ as too rich to admit a useful exchangeable structure, but may be willing to assume exchangeability on a simpler submodel S.

Third, we alter the group of permutations to model subpopulations with different stochastic characteristics. To motivate this, consider an econometrician studying the impact of education on the economic success of an individual. Each observation embodies information about an individual's characteristics such as parents' educational level, number of years at school, average income during the first 5 years after college, number of job offers received, and so on. These variables are represented by a polish space Y.

The econometrician's model divides the population into $\{1, \ldots, K\}$ disjoint subpopulations and assumes exchangeability *conditional* on each subpopulation. For instance, with two subpopulations, men and women, the econometrician assumes similarity of the impact of education on economic performance when restricted to samples of individual of the same gender, but he is agnostic about any statistical relationship across genders.

To incorporate conditional exchangeability in our framework, we let the outcome space be $S \equiv \{1, \ldots, K\} \times Y$. A sequence of observations is now of the form $\big((k_1, y_1), (k_2, y_2), \ldots\big)$ where (k_n, y_n) indicates that the nth observation belongs to subpopulation k_n. Acts and preferences are defined in the usual way.

The decision maker may not view the experiments to be exchangeable since he may not necessarily assesses an event $A \subset Y$ as equally likely in all experiments, regardless of the subpopulation the individual was drawn from. The decision maker may also find it unreasonable to require exchangeability with respect to all events in S. Such a requirement would commit him, for example, to the belief that each subpopulation appears in a well-defined limiting frequency in almost all samples. The econometrician's belief in an invariant structure for each subpopulation seems quite distinct from whether or not he should also believe that the subpopulations are sampled in an i.i.d. manner.

In light of the above, Corollary 1 cannot be applied. However we can apply Theorem 1 with a modified transformation. For each subpopulation k, let ρ_k be the stopping time that selects the next time in which the observation belongs to subpopulation k. Assume that the set of sequences in which each subpopulation occurs infinitely often is the complement of a null set. This ensures that the stopping time ρ_k is finite outside a null set.

Restrict attention to acts that depend on the first coordinate only. Write any such act as a vector $f = (f_1, \ldots, f_K)$ that pays $f_k(y)$ if the individual belongs to subpopulation k and has characteristics y. Define the transformation:

$$\tau_k(\omega) \equiv T^{\rho_k(\omega)}(\omega)$$

where T is the shift transformation. Thus, τ_k is nothing but the shift applied as many times as needed to get to the next observation in subpopulation k. Applying Theorem 1 one k at a time yields the empirical limit $f^\star = (f_1^\star, \ldots, f_K^\star)$ and the conclusion $f \sim f^\star$. This example illustrates the power of the general subjective ergodic theorem in handling non-trivial variations on the notion of similarity.[34]

To conclude, we note that a comprehensive theory of partial exchangeability is beyond the scope of this paper. What we tried to do above is to present simple, informal examples to illustrate two related ideas. First, that similarity, broadly construed, need not be tied with the notion of invariance relative to permutations. Although such invariance (*i.e.,* exchangeability) is an important base-line case and the focus of the present paper, it does not

[34] In the process we have glossed over important details to get our basic point across.

exhaust the richness of the idea of similarity. Second, subjective ergodic theory provides a general framework to study the role of similarity in decision making.

7 Discussion

7.1 De Finetti's View of Similarity as the Basis for Probability Judgement

In his classic 1937 article,[35] de Finetti wondered how an insurance company may evaluate the probability that an individual dies in a given year. To evaluate this probability one must first choose, in de Finetti's terms, a class of "similar" events, then use the frequency as base-line estimate of the probability. For example, one may consider as similar the event: "death in a given year of an individual of the same age [...] and living in the same country." De Finetti notes that the choice of a class of "similar" events is, in itself, subjective,[36] since one could have easily considered "not individuals of the same age and country, but those of the same profession and town, ... etc, where one can find a sense of 'similarity' that is also plausible."

In this paper we have assumed, like the rest of the literature, a fixed experiment S and a relationship of exchangeability linking repetitions of this experiment. A more abstract, but potentially more useful, way to think about exchangeability is to view it as reflecting the decision maker's subjective "theory" of those aspects of the underlying state space that he views as similar. Here, we interpret "similar experiments" as ones the decision maker subjectively views as governed by a stable stochastic structure.

To make this formal, think of a sequence of experiments as mappings $O_t : \tilde{\Omega} \to \tilde{S},\ t = 1, \ldots$, from an underlying state space $\tilde{\Omega}$ to an abstract set of labels $\tilde{S}$. Here, we interpret the observation $O_t(\tilde{\omega})$ as the label given to the measurement made at time t. In de Finetti's example above, an observation may be of the form "a man of a given age group, profession and town died."

A decision maker's *exchangeability structure* $\mathcal{O}$ is a sequence of observa-

[35] All references are to the original text in French, pages 20-22.
[36] In the original text, de Finetti uses the term *arbitraire*.

tions $\{O_t\}$ which he subjectively views as exchangeable. This can be viewed as the decision maker's theory of the world in the sense that it decomposes the uncertainty he faces into a stationary part that can be estimated from frequencies, and an idiosyncratic noise. Another decision maker may adopt a distinct structure $\mathcal{O}'$ with a different decomposition of uncertainty. In this case, what appears as idiosyncratic noise to one decision maker may be viewed as predictable by another.

An important question is whether different exchangeability structures can be combined to form a structure with more narrowly defined events. De Finetti seemed to have anticipated this issue, noting that the prevision of probabilities will "in general be more difficult the narrower the class of events considered." Note that this statement makes little sense for an expected utility decision maker. Al-Najjar (2009) provides a formal model based on classical statistics explaining that refining exchangeability structures exacerbates the problem of over-fitting when data is scarce.

7.2 A Subjectivist Interpretation of Classical Statistics

Kreps (1988) argued that de Finetti's theorem is the fundamental theorem underlying all of statistical inference. But since de Finetti's theorem assumes that the statistician has probabilistic beliefs over parameters, this is clearly a non-starter for the purpose of understanding, let alone reconciling, the subjectivist view with the prevailing classical practice in statistics and econometrics. Our framework neither favors sides, nor does it claim to resolve the decades-old debate between classical and Bayesian statistics. Rather, by disentangling subjectivity and exchangeability from Bayesianism, we can shed some light on the sources of disagreement.

A common misconception associates this disagreement with the classicists' use of concepts like "true parameters" which subjectivists like de Finetti find meaningless. Within our framework, all decision maker's are subjectivists! The difference between them is how they resolve uncertainty about the parameters. References to "true" parameters, although confusing as a rhetorical device, can be given a rigorous purely subjectivist foundation using the concept of exchangeability. Classical statistical models and estimation procedures can, in principle, be derived from preference with parameters and implied probabilities that are purely subjective. The classicist, however, is unwilling to commit to the Bayesian inductive principle of forming a subjective belief on the parameters and incorporate new evidence using Bayesian updating. The classicist opts instead to draw finite sample inferences using procedures that require robustness or uniformity across parameters.

7.3 Learning and Predictions

Exchangeability and de Finetti's theorem are closely connected to learnability. Jackson, Kalai, and Smorodinsky (1999) characterize stochastic processes P that admit a decomposition: $P = \int P^\theta \, d\mu$, where the parameters are 'fine enough' to be predictive, yet not so fine to be unlearnable (we refer the reader to their paper for formal definitions and motivation). In the special case of an exchangeable process P, their results characterize de Finetti's representation as the unique learnable and predictive decomposition of P. As data accumulates, learning the true i.i.d. component P^θ is the best a Bayesian learner can do.

Jackson, Kalai, and Smorodinsky (1999) assume Bayesian beliefs and updating, an assumption that rules out preference that are incomplete and/or probabilistically unsophisticated. For such preferences, non-Bayesian learning procedures are possible. Gray and Davisson (1974, Theorem 4.2), for instance, provide general results for the limiting behavior of frequentist learning procedures in a context that fits ours.

A Proofs

A.1 Regularity

The body of the paper restricts attention to finitely-based acts. The results hold more generally, requiring only the regularity condition below (which hold trivially when acts are finitely-based).

Given a sequence of transformations $\phi^n : \Omega \to \Omega$ and a transformation $\phi : \Omega \to \Omega$, we write $\phi^n \to \phi$ if for every k, there exists n_k such that the first k coordinates of ϕ^n agree with the first k coordinates of ϕ, for all $n \geq n_k$. Note that this is a strong notion of convergence. It is used in the following assumption, which requires that the convergence of transformations imply the convergence of acts:

Assumption 7 (Regularity) *Consider an act f, a transformation $\phi : \Omega \to \Omega$ and a sequence of transformations $\phi^n : \Omega \to \Omega$. If $\phi^n \to \phi$ then $f \circ \phi^n \to f \circ \phi$.*

Since $\phi^n \to \phi$ is a strong premise, the above assumption is mild. Observe also that Assumption 7 is trivially satisfied for every finitely based acts.

Lemma A.1 *Exchangeability and Assumption 7 imply that 2.*

Proof: For each $j = 0, 1, 2, ...$ fixed, define the following sequence of permutations:

$$\phi^{j,n}(s_1, s_2, ...) = (s_{j+1}, s_{j+2}, ..., s_n, s_1, s_2, ..., s_j, s_{n+1}, ...).$$

It is easy to see that $\phi^{j,n} \to T^j$, where T is the shift, T^0 is the identity and $T^j = T \circ ... \circ T$ (j times). Given $f \in \mathcal{F}$, define $f^{i,n} \equiv \frac{1}{i}\sum_{j=0}^{i-1} f \circ \phi^{j,n}$ and $f^i \equiv \frac{1}{i}\sum_{j=0}^{i-1} f \circ \tau^j$. By exchangeability, $f \sim f^{i,n}$. By Assumption 7, $f \circ \phi^{j,n} \to f \circ T^j$. It is clear that

$$\{\omega : \lim_{n\to\infty} f^{i,n}(\omega) \neq f^i(\omega)\} \subset \cup_{j=0}^{i-1}\{\omega : \lim_{n\to\infty} f^{i,n} \circ \phi^{i,n}(\omega) \neq f \circ T^j(\omega)\}.$$

Lemma A.2 below implies that union in the right above is null and, since subsets of null sets are null (see Lemma A.3 below), $f^{i,n} \to f^i$. By continuity (Assumption 3), $f \sim f^i$, as we wanted to show. ■

Lemma A.2 *Let C_n be null for all $n \in \mathbb{N}$. Then, $C = \cup_{n \in \mathbb{N}} C_n$ is null.*

Proof: Let f, g, h be arbitrary acts. Define $A^N \equiv \cup_{n=1}^N C_n$; $f^N \equiv f 1_{A^N} + h 1_{(A^N)^c}$, and $g^N \equiv g 1_{A^N} + h 1_{(A^N)^c}$. Observe that $f^N = \sum_{n=1}^N f 1_{C_n} + h 1_{(A^N)^c}$ and $g^N = \sum_{n=1}^N g 1_{C_n} + h 1_{(A^N)^c}$. Using the nullness of C_n for each $n = 1, 2, ...$, we have:

$$\begin{array}{rcl} f^N & = & f 1_{C_1} + f 1_{C_2} + \cdots + f 1_{C_N} + h 1_{(A^N)^c} \\ & \sim & g 1_{C_1} + f 1_{C_2} + f 1_{C_3} + \cdots + f 1_{C_N} + h 1_{(A^N)^c} \\ & \sim & g 1_{C_1} + g 1_{C_2} + f 1_{C_3} + \cdots + f 1_{C_N} + h 1_{(A^N)^c} \\ & & \cdots \\ & \sim & g 1_{C_1} + g 1_{C_2} + \cdots + g 1_{C_N} + h 1_{(A^N)^c} \\ & = & g^N. \end{array}$$

Note that the set $\{\omega \in \Omega : \lim_{N \to \infty} f^N(\omega) \neq (f 1_C + h 1_{C^c})(\omega)\}$ is empty. It is easy to see that the fact that the preference is reflexive (Assumption 1) implies that the empty set is $\succcurlyeq$-null. Therefore, $f^N \to f 1_C + h 1_{C^c}$. Similarly, $g^N \to g 1_C + h 1_{C^c}$. Define $b(\omega) \equiv \max\{|f(\omega)|, |g(\omega)|, |h(\omega)|\}$. Since $|f^N|, |g^N| < b$, $f^N \sim g^N$ for all N, and $f^N \to f 1_C + h 1_{C^c}$ and $g^N \to g 1_C + h 1_{C^c}$, continuity (Assumption 3) implies $f 1_C + h 1_{C^c} \sim g 1_C + h 1_{C^c}$. Since f, g, h are arbitrary, C is null. ■

Lemma A.3 *If $A \subset E$ and E is $\succcurlyeq$-null then A is $\succcurlyeq$-null.*

Proof: Let $f, g, h \in \mathcal{F}$. Define $f' \equiv f 1_A + h 1_{E \setminus A}$ and $g' \equiv g 1_A + h 1_{E \setminus A}$. Since E is null, we have:

$$\left[\begin{array}{ll} f'(\omega), & \text{if } \omega \in E \\ h(\omega), & \text{if } \omega \notin E \end{array} \right] \sim \left[\begin{array}{ll} g'(\omega), & \text{if } \omega \in E \\ h(\omega), & \text{if } \omega \notin E \end{array} \right] \tag{10}$$

Note, however that the left and right side above are respectively:

$$\left[\begin{array}{ll} f(\omega), & \text{if } \omega \in A \\ h(\omega), & \text{if } \omega \notin A \end{array} \right] \text{ and } \left[\begin{array}{ll} g(\omega), & \text{if } \omega \in A \\ h(\omega), & \text{if } \omega \notin A \end{array} \right]. \tag{11}$$

Therefore, A is $\succcurlyeq$-null. ■

A.2 Proof of Theorem 1

In this section, we prove Theorem 1 and make comments (in footnotes) how to adapt this proof to the case of Corollary 1, where the result is stated for finitely based acts and without requiring Assumption 7.[37]

We begin by assuming that (2) holds. For each $f \in \mathcal{F}$, define:

$$\bar{f}(\omega) \equiv \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right), \tag{12}$$

and

$$\underline{f}(\omega) \equiv \liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right). \tag{13}$$

It is easy to see that if $f \in \mathcal{F}$, then $\bar{f}$ and $\underline{f} \in \mathcal{F}$. We have the following:

Lemma A.4 *For all $\omega \in \Omega$, $\bar{f}(\tau\omega) = \bar{f}(\omega)$ and $\underline{f}(\tau\omega) = \underline{f}(\omega)$.*

Proof: Fix $\omega \in \Omega$. We have:

$$\begin{aligned} \bar{f}(\tau\omega) &= \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \tau\omega\right) \\ &= \limsup_{n\to\infty} \left[\frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right) + \frac{f(\tau^n\omega) - f(\omega)}{n} \right]. \end{aligned}$$

Since f is bounded, the above limit is equal to:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right) = \bar{f}(\omega),$$

which concludes the proof for $\bar{f}$. The proof for $\underline{f}$ is analogous. ■

[37]Observe that for finitely based acts, Assumption 7 holds trivially, so that (2) is satisfied for all finitely based acts if Assumption 4 holds.

Lemma A.5 *Define*

$$A \equiv \left\{ \omega \in \Omega : \bar{f}(\omega) > a \right\},$$

for some $a \in \mathbb{R}$. Then, A is τ-invariant, that is, $\omega \in A \Leftrightarrow \tau\omega \in A$. Moreover, for all $\omega \in A$ and $n = 0, 1, 2, ...$ there exists a least integer $m(n, \omega) \geq n$ such that

$$\frac{1}{m(n,\omega) - n + 1} \sum_{j=n}^{m(n,\omega)} f(\tau^j \omega) > a.$$

Proof: The invariance of A follows directly from the previous result: $\omega \in A \Leftrightarrow \bar{f}(\tau\omega) = \bar{f}(\omega) > a \Leftrightarrow \tau\omega \in A$. Fix $\omega \in A$ and let $\varepsilon \equiv \frac{\bar{f}(\omega)-a}{2} > 0$, that is, $\bar{f}(\omega) = a + 2\varepsilon$. For a contradiction, assume that there is a n such that for all $m \geq n$,

$$\sum_{j=n}^{m} f(\tau^j \omega) \leqslant a(m - n + 1).$$

Let $L = \sum_{j=0}^{n-1} f(\tau^j \omega)$. By the definition of $\bar{f}(\omega)$, there is a $m > \max\left\{ \frac{L+a-an}{\varepsilon}, n \right\}$ such that:

$$m(a + \varepsilon) < \sum_{j=0}^{m} f(\tau^j \omega) = \sum_{j=0}^{n-1} f(\tau^j \omega) + \sum_{j=n}^{m} f(\tau^j \omega) \leqslant L + a(m - n + 1).$$

This means that $m\varepsilon < L - an + a$, contradicting the choice of m. ■

The following result generalizes the Maximal Ergodic Theorem (Billingsley (1965, p. 25)):

Proposition A.6 *Define*

$$A \equiv \left\{ \omega \in \Omega : \bar{f}(\omega) > a \right\},$$

for some $a \in \mathbb{R}$. Then, for any $h \in \mathcal{F}$, $fAh \succcurlyeq aAh$.

Proof: From Lemma A.5, for each $\omega \in A$, there is a least integer $m(\omega) = m(0, \omega) \geq 0$ such that $\sum_{j=0}^{m(\omega)} f(\tau^j \omega) > a(m(\omega) + 1)$. For each $l = 0, 1, \ldots,$ define:

$$A_l \equiv \{\omega \in A : m(\omega) = l\}.$$

Note that A_l is measurable, since $\sum_{j=0}^{l} f(\tau^j\omega)$ is measurable for every l. It is clear that $A = \cup_{l=0}^{\infty} A_l$.[38] For each $L \in \mathbb{N}$, define: $E_L \equiv \cup_{l=0}^{L-1} A_l$.[39] Observe that A_l, E_L and A are all stopping-time sets. Now, define $f^a(\omega) \equiv f(\omega) - a$. Observe that for all j and $\omega \in \Omega$,

$$f^a(\tau^j\omega)1_{E_L}(\tau^j\omega) \geq f^a(\tau^j\omega). \tag{14}$$

Indeed, if $\tau^j\omega \in E_L$, then the inequality is obvious. If $\tau^j\omega \notin E_L$, then $\tau^j\omega \notin A_0$, that is, $f(\tau^j\omega) \leq a$ or, equivalently, $f^a(\tau^j\omega) \leq 0 = f^a(\tau^j\omega)1_{E_L}(\tau^j\omega)$.

Now we will define, for each $\omega \in \Omega$, a number $\ell(\omega) \in \{1, ..., L\}$ such that:

$$\sum_{j=0}^{\ell(\omega)-1} f^a(\tau^j\omega)1_{E_L}(\tau^j\omega) \geq 0. \tag{15}$$

If $\omega \in E_L$, there exists $m(\omega) \leqslant L-1$ such that $\sum_{j=0}^{m(\omega)} f(\tau^j\omega) > a(m(\omega)+1)$, that is, $\sum_{j=0}^{m(\omega)} f^a(\tau^j\omega) \geq 0$. In this case, we can just take $\ell(\omega) = m(\omega)+1 \leqslant L$. Indeed, (14) implies:

$$\sum_{j=0}^{\ell(\omega)-1} f^a(\tau^j\omega)1_{E_L}(\tau^j\omega) \geq \sum_{j=0}^{\ell(\omega)-1} f^a(\tau^j\omega) \geq 0.$$

Now, if $\omega \notin E_L$, we choose $\ell(\omega) = 1$ (observe that the inequality in (15) is satisfied because $1_{E_L}(\omega) = 0$).

Now, fix $\omega \in \Omega$ and take any $N \in \mathbb{N}$ greater than $L > 1$. Put $n_0 = 0$ and define recursively: $n_{k+1} = n_k + \ell(\tau^{n_k}\omega)$, until finding K such that: $N - 1 \geq n_K \geq N - L > n_{K-1}$. This is possible because $\ell(\omega) \leq L$ for all $\omega \in \Omega$. Thus,

$$\begin{aligned}
\sum_{j=0}^{N-1} f^a\left(\tau^j\omega\right)1_{E_L}(\tau^j\omega) &= \sum_{k=1}^{K}\sum_{j=n_{k-1}}^{n_k-1} f^a\left(\tau^j\omega\right)1_{E_L}(\tau^j\omega) \\
&\qquad + \sum_{j=n_K}^{N-1} f^a\left(\tau^j\omega\right)1_{E_L}(\tau^j\omega) \\
&\geq 0 + \sum_{j=n_K}^{N-1}(-B),
\end{aligned}$$

[38] In the finitely-based case, A_l (or more precisely, the indicator function 1_{A_l}) is finitely-based for each l, but not A. We will not apply (2) for functions depending on A.

[39] In the finitely-based case, E_L is finitely-based because each A_l is.

from (15), where $B>0$ is chosen such that $f^a(\omega)=f(\omega)-a>-B$ for every $\omega\in\Omega$. Since $N-n_K\leq L$, we conclude that

$$\sum_{j=0}^{N-1} f^a\left(\tau^j\omega\right) 1_{E_L}(\tau^j\omega)\geq -LB. \tag{16}$$

Define

$$\begin{aligned}
f_L(\omega) &\equiv f(\omega)\, 1_{E_L}(\omega)+h(\omega)1_{A^c}(\omega);\\
f_L^N(\omega) &\equiv \frac{1}{N}\sum_{j=0}^{N-1} f_L\left(\tau^j\omega\right);\\
h^N(\omega) &\equiv \frac{1}{N}\sum_{j=0}^{N-1} h(\tau^j(\omega))1_{A^c}(\tau^j(\omega));\\
g_L(\omega) &\equiv a1_{E_L}(\omega)+h(\omega)1_{A^c}(\omega);\\
g_L^N(\omega) &\equiv \frac{1}{N}\sum_{j=0}^{N-1} g_L(\tau^j\omega);\\
g_{L,M}(\omega) &\equiv a1_{E_L}(\omega)+h(\omega)1_{A^c}(\omega)-\frac{LB}{M};\\
g_{L,M}^N(\omega) &\equiv \frac{1}{N}\sum_{j=0}^{N-1} g_{L,M}(\tau^j\omega);
\end{aligned}$$

Then, using (16), we have for all $\omega\in\Omega$:

$$\begin{aligned}
f_L^N(\omega) &= \frac{1}{N}\sum_{j=0}^{N-1}\left[f^a\left(\tau^j\omega\right)+a\right] 1_{E_L}(\tau^j\omega)+h^N(\omega)\\
&= \frac{1}{N}\sum_{j=0}^{N-1} a1_{E_L}(\tau^j\omega)+\frac{1}{N}\sum_{j=0}^{N-1} f^a\left(\tau^j\omega\right) 1_{E_L}(\tau^j\omega)+h^N(\omega)\\
&\geq g_L^N(\omega)+h^N(\omega)-\frac{LB}{N}=g_{L,N}^N(\omega).
\end{aligned}$$

By monotonicity (assumption 2), $f_L^N\succcurlyeq g_{L,N}^N$ and by (2), $f_L\sim f_L^N$ and $g_{L,N}^N\sim g_{L,N}$.[40] Therefore transitivity gives $f_L\succcurlyeq g_{L,N}$. Let $\{f_L,f_L,...\}$ and

[40] Note that in the finitely-based case, these functions will always be finitely-based and, therefore, (2) holds.

$\{g_{L,L+1}, g_{L,L+2}, ..., g_{L,N}, ...\}$ be sequences of functions indexed by N. It is clear that $f_L \to f_L$ and $g_{L,N} \to g_L$ when $N \to \infty$. By continuity (Assumption 3), we have $f_L \succcurlyeq g_L$. Now, when $L \to \infty$, we have $f_L \to fAh$ and $g_L \to aAh$. Therefore, again by continuity, $fAh \succcurlyeq aAh$, as we wanted to show. ■

Corollary A.7 *Fix an act f and $b \in \mathbb{R}$ and define*

$$B \equiv \left\{\omega \in \Omega : \underline{f}(\omega) < b\right\}.$$

Then, for any $h \in \mathcal{F}$, $bBh \succcurlyeq fBh$.

Proof: For this, it is sufficient to repeat the arguments in the proof of proposition A.6, with the obvious adaptations. ■

Corollary A.8 *Fix an act f, $a, b \in \mathbb{R}$ and define A and B as in the previous two results. Then, for any stopping-time set C and any $h \in \mathcal{F}$, we have: $f(A \cap C)h \succcurlyeq a(A \cap C)h$ and $b(B \cap C)h \succcurlyeq f(B \cap C)h$.*

Proof: Repeat the proofs of Proposition A.6 and Corollary A.7 substituting f by $f1_C$ and A (B) by $A \cap C$ $(B \cap C)$ whenever appropriate. ■

Now, we can complete the proof of the theorem with the following lemmas:

Lemma A.9 *Let us define $f^n(\omega) = \frac{1}{n}\sum_{j=0}^{n-1} f(\tau^j\omega)$. Then, $f^n \to f^\star$ and $f \sim f^\star$.*

Proof: Fix $b < a$ and define the set

$$C \equiv \left\{\omega \in \Omega : \liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f\left(\tau^j\omega\right) < b < a < \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f\left(\tau^j\omega\right)\right\}.$$

It is clear that C is τ-invariant and that $C = A \cap B$, where A and B are the sets defined in Proposition A.6 and Corollary A.7. By Corollary A.8, $bCh \succcurlyeq aCh$, for every $h \in \mathcal{F}$. Since $a > b$, monotonicity (Assumption 2) implies that $aCh \sim bCh$. Let $x, y \in \mathbb{R}$, $x < y$ and find $\alpha, \beta \in \mathbb{R}$, $\beta > 0$ such that $x = \beta b + \alpha$ and $y = \beta a + \alpha$. Now define the function $\tilde{f}(\omega) = \beta f(\omega) + \alpha$. Then,

$$C \equiv \left\{ \omega \in \Omega : \liminf_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{f}\left(\tau^j \omega\right) < x < y < \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{f}\left(\tau^j \omega\right) \right\}.$$

Then, for any $x, y \in \mathbb{R}$, we can , we have $xCh \sim yCh$. Now assume that C is not null. Then there are functions g and $h \in \mathcal{F}$ such that $gCh \succ h$. Since g and h are bounded, let $x \equiv \sup_{\omega \in C} g(\omega)$ and $y \equiv \inf_{\omega \notin C} h(\omega)$. By monotonicity, $xCh \succcurlyeq gCh$ and $h \succcurlyeq yCh$. This implies that $gCh \sim h$, in contradiction to $gCh \succ h$. This shows that C is null.

Now, the set where the limit does not exist is formed by the countable union of sets of the above form (it is sufficient to take a and b in the rationals). Thus, the existence of the limit except in a null set comes from the fact that a countable union of null sets is null (see Lemma A.2). This shows that $f^n \to f^\star$. By (2), $f^n \sim f$ for all $n \in \mathbb{N}$ and by continuity (Assumption 3), $f^\star \sim f$. ∎

Lemma A.10 *$f^\star$ is τ-invariant.*

Proof: Let E be the set of ω for which $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right)$ exists. Repeating the argument of Lemma A.4, we have for $\omega \in E$:

$$\begin{aligned} f^\star(\tau\omega) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \tau \omega\right) \\ &= \lim_{n\to\infty} \left[\frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right) + \frac{f(\tau^n \omega) - f(\omega)}{n} \right]. \end{aligned}$$

Since f is bounded, the above limit is equal to:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right) = f^\star(\omega),$$

which shows that E is τ-invariant and that $f^\star \tau = f^\star$ in E. For $\omega \notin E$, $\tau\omega \notin E$ and $f^\star(\omega) = 0 = f^\star(\tau\omega)$. Therefore, $f^\star \tau = f^\star$ everywhere. ∎

Lemma A.11 *If $\succcurlyeq$ is τ-ergodic, then $f^\star$ is constant except in a null set.*

Proof: For every $a \in \mathbb{R}$, the set $A = \{\omega : f^\star(\omega) > a\}$ is τ-invariant, because $f^\star\tau = f^\star$. If $\succcurlyeq$ is τ-ergodic, then A is trivial. Since this is valid for all $a \in \mathbb{R}$, then $f^\star$ is constant except in a null set, which concludes the proof. ■

Finally, the converse implication in Theorem 1 is proved by the following:

Lemma A.12 *If the limit*

$$f^\star(\omega) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j \omega\right)$$

exists except on a $\succcurlyeq$-null event for all $f \in \mathcal{F}$ and $f \sim f^\star$, then (2) holds for all $f \in \mathcal{F}$.

Proof: Assume that there exists $f \in \mathcal{F}$ and $n \in \mathbb{N}$ such that $f \not\sim g \equiv \frac{1}{n}\sum_{i=0}^{n-1} f \circ \tau^j$. Then,

$$\begin{aligned} g^\star(\omega) &= \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} \left(\frac{1}{n} \sum_{i=0}^{n-1} f \circ \tau^i \right) \circ \tau^j \\ &= \frac{1}{n} \lim_{N\to\infty} \sum_{i=0}^{n-1} \left(\frac{1}{N} \sum_{j=0}^{N-1} f \circ \tau^{i+j} \right). \end{aligned}$$

The idea of the proof of Lemma A.4 can be used to show that the limit exists for f iff it exists for $f \circ \tau^i$ and they are equal. Thus, the above is:

$$\frac{1}{n} \sum_{i=0}^{n-1} \lim_{N\to\infty} \left[\frac{1}{N} \sum_{j=0}^{N-1} \left(f \circ \tau^i\right) \circ \tau^j \right] = \frac{1}{n} \sum_{i=0}^{n-1} \left(f \circ \tau^i\right)^\star(\omega) = \frac{1}{n} \sum_{i=0}^{n-1} f^\star(\omega) = f^\star(\omega).$$

In other words, $f^\star(\omega) = g^\star(\omega)$, off a null set. By monotonicity and the definition of null sets, $f \sim g$, which is a contradiction. ■

A.3 Proof of Theorem 2

The proof of Theorem 2 is based in the Ergodic Decomposition Theorem, which we state here for reader's convenience. A proof of this theorem can be found in Varadarajan (1963), section 4. See also Löh (2006) or Viana (2008) for a more accessible discussion and motivation.

Recall that Υ is the set of exchangeable probability measures over (Ω, Σ). An event $A \in \Sigma$ is exchangeable if $\pi A = A$ for all $\pi \in \Pi$ and P-trivial if either $P(A) = 0$ or $P(A) = 1$. The set of exchangeable (or symmetric) events is denoted by $\mathcal{S}$.

A measure $P \in \Upsilon$ is ergodic if all exchangeable events are P-trivial, that is, $A \in \mathcal{S} \Rightarrow P(A) = 0$ or 1. Let $\mathcal{E} \subset \Upsilon$ denote the set of ergodic exchangeable probability measures.

Ergodic Decomposition Theorem. *The set $\mathcal{E}$ is non-empty and there exists a map $\beta : \Omega \to \mathcal{E}$ with the following properties:*

1. *If $A \in \Sigma$, then the map $\omega \mapsto \beta(\omega, A)$ is measurable, where $\beta(\omega, A)$ denotes the value of the measure $\beta(\omega)$ at the set A.*

2. *For all $\pi \in \Pi$, $\beta(\omega) = \beta(\pi\omega)$.*

3. *If $\nu \in \mathcal{E}$, define $\Omega_\nu = \{\omega \in \Omega : \beta(\omega) = \nu\}$. Then, $\Omega_\nu \in \Sigma$, $\nu(\Omega_\nu) = 1$ and ν is the unique ergodic symmetric measure with support in Ω_ν.*

4. *For each $\nu \in \mathcal{E}$ and $f \in \mathcal{F}$, $\int_{\Omega_\nu} f d\nu = \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} f(\tau^j \omega)$ for ν-almost all $\omega \in \Omega_\nu$.*

5. *For all $P \in \Upsilon$ and all $A \in \Sigma$,*

$$P(A) = \int_\Omega \beta(\omega, A) dP(\omega).$$

Moreover, this map β is essentially unique, in the sense that if β' satisfies the above properties, then $P(\{\omega : \beta(\omega) \neq \beta'(\omega)\}) = 0$ for all $P \in \Upsilon$.

Also, there is a one-to-one measurable map $P \mapsto \sigma_P$ from Υ to the set of all measures on $\mathcal{E}$ such that, for all $A \in \Sigma$ and $P \in \Upsilon$,

$$P(A) = \int_{\mathcal{E}} \nu(A) d\mu_P(\nu).$$

Proof of Theorem 2: Let Θ be the set of probability measures in S. For each $\theta \in \Theta$, let P^θ denote the product measure in Ω obtained from θ. The set $\{P^\theta : \theta \in \Theta\}$ is the set of extreme points of the set of invariant measures (see Hewitt and Savage (1955), Theorem 5.3, p. 478). (This also coincides with the set of ergodic measures in Ω.) Fix the map β given by the Ergodic decomposition Theorem. Using the notation in that theorem, define $E^\theta \equiv \Omega_{P^\theta}$. By (3) in that theorem, $P^\theta(E^\theta) = 1$ and P^θ is the unique exchangeable probability measure with support in E^θ. By (4), $f^\star(\omega) = \int_{E^\theta} f dP^\theta$ for P^θ-almost all $\omega \in E^\theta$. This concludes the proof. ■

A.4 Proof of Theorem 3

Consider the sup-norm in $\mathcal{F}$:

$$\|f\| = \sup_{\omega \in \Omega} |f(\omega)|.$$

A subset $\tilde{\mathcal{F}} \subset \mathcal{F}$ is separable if there exists a countable dense subset, that is, a countable set $H \subset \tilde{\mathcal{F}}$ such that for every $f \in \tilde{F}$ and $\epsilon > 0$, there exists $h \in H$ such that $\|f - h\| < \epsilon$. We need the following:

Lemma A.13 *Let S be a Polish space and $\mathcal{S}$, its Borel σ-algebra. Then, there exists a countable algebra $\mathcal{S}^\circ$ of subsets of S that generates $\mathcal{S}$.*

Proof: Since S is Polish, by Royden (1968, Theorem 8, p. 326) it is Borel isomorphic to (i) $[0,1]$; (ii) $\mathbb{N}$; or (iii) a finite set. Sets A and B are Borel isomorphic if there is a measurable bijective map $h : A \to B$, with measurable inverse. Consider first the case where S is Borel isomorphic to $[0,1]$. Let $\mathcal{A}$ denote the collection of finite unions of intervals of the form $[a,b)$, for $a, b \in \mathbb{Q} \cap [0,1]$. It is easy to see that $\mathcal{A}$ is a countable algebra that generates the Borel field of $[0,1]$. Since intersections and set difference is preserved under the inverse of a function, then $\mathcal{S}^\circ \equiv h^{-1}(\mathcal{C})$ is also a countable algebra. Since $h^{-1}(\sigma(\mathcal{C})) = \sigma(h^{-1}(\mathcal{C}))$, then $\mathcal{S}^\circ$ generates the Borel field $\mathcal{S}$.

If S is Borel isomorphic to $\mathbb{N}$, take as $\mathcal{A}$ the algebra of all singletons of $\mathbb{N}$, which generates its Borel field. Then, repeat the ideas above. Finally, if S is Borel isomorphic to a finite set, take as $\mathcal{A}$ the power set of this finite set and repeat the same arguments. ■

Let $\mathcal{S}^1$ denote the class of sets of the form $A \times S \times S \times ...$, for $A \in \mathcal{S}^\circ$. It is clear that $\mathcal{S}^1$ is also a countable algebra. Following Dunford and Schwartz (1958), IV.2.12, p. 240, let $B(\Omega, \mathcal{S}^1)$ denote the space of all uniform limits of finite linear combinations of characteristic functions of sets in $\mathcal{S}^1$.

Lemma A.14 *$B(\Omega, \mathcal{S}^1)$ is separable.*

Proof: The set $B(\Omega, \mathcal{S}^1)$ is the closed linear span of a countable set, namely, $\cup_{n\in\mathbb{N}} \{\sum_{i=1}^{n} \alpha_i 1_{A_i} : \alpha_i \in \mathbb{Q}; A_i \in \mathcal{S}^1, \text{ for } i = 1, 2, ..., n\}$. Therefore, it is separable. ■

Lemma A.15 *Let $\tilde{\mathcal{F}} \subset \mathcal{F}$ be separable. Then, there exists $\Omega' \subset \Omega$ with $\succcurlyeq$-null complement, such that the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} f\left(T^j \omega\right)$$

exists for all $f \in \tilde{\mathcal{F}}$ and $\omega \in \Omega'$.

Proof: Let $H = \{h_n : n \in \mathbb{N}\}$ be a countable dense set of $\tilde{\mathcal{F}}$. By Theorem 1, there exist a set Ω_n such that $\Omega \setminus \Omega_n$ is $\succcurlyeq$-null, and the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} h_n\left(T^j \omega\right)$$

exists for all $\omega \in \Omega^n$. Define $\Omega' \equiv \cap_n \Omega^n$ and, for each $f \in \tilde{\mathcal{F}}$ and $N \in \mathbb{N}$, $f^N(\omega) \equiv \frac{1}{N}\sum_{j=0}^{N-1} f\left(T^j\omega\right)$. By Lemma A.2, $\Omega \setminus \Omega'$ is $\succcurlyeq$-null. Therefore, it is sufficient to show that for all $\omega \in \Omega'$ there exists $\lim_{N\to\infty} f^N(\omega)$ or, equivalently, that $\{f^N(\omega)\}$ is Cauchy. Given $\varepsilon > 0$, choose h_n such that $\|f - h_n\| < \frac{\varepsilon}{3}$, which can be done because $\{h_n\}$ is dense. Now, choose n_ϵ

such that $N, M > n_\epsilon$ implies $|\, h_n^N(\omega) - h_n^M(\omega)\,| < \frac{\varepsilon}{3}$. Therefore:

$$\begin{aligned}
|\, f^N(\omega) - f^M(\omega)\,| &\leq |\, f^N(\omega) - h_n^N(\omega)\,| + |\, h_n^N(\omega) - h_n^M(\omega)\,| \\
&\qquad\qquad + |\, h_n^M(\omega) - f^M(\omega)\,| \\
&\leq \frac{1}{N}\sum_{j=0}^{N-1} |\, f\left(T^j\omega\right) - h_n(T^j\omega)\,| + \frac{\epsilon}{3} \\
&\qquad\qquad + \frac{1}{M}\sum_{j=0}^{M-1} |\, f\left(T^j\omega\right) - h_n(T^j\omega)\,| \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,
\end{aligned}$$

that is, $\{f^N(\omega)\}$ is Cauchy for all $\omega \in \Omega'$. This shows that the limit $\lim_{N\to\infty} f^N(\omega)$ exists for all $\omega \in \Omega'$ and $f \in \tilde{\mathcal{F}}$. ■

Lemma A.16 *Let $\tilde{\mathcal{F}} \subset \mathcal{F}$ be separable. If $\succcurlyeq$ is ergodic, then there exists Ω' so that the limit $f^\star(\omega)$ exists for all $\omega \in \Omega'$ and $f^\star$ is constant in Ω', for all $f \in \tilde{\mathcal{F}}$.*

Proof: As in the proof of Lemma A.15, let $H = \{h_n : n \in \mathbb{N}\}$ be a countable dense set of $\tilde{\mathcal{F}}$. Differently from before, let Ω_n be a set with $\succcurlyeq$-null complement, such that the limit exists for h_n and $h_n^\star$ is constant in Ω_n. This set exists by the last part of the Subjective Ergodic Theorem (Theorem 1). As before, define $\Omega' \equiv \cap_n \Omega^n$, whose complement is $\succcurlyeq$-null by Lemma A.2.

From Lemma A.15, we know that the limit $\lim_{N\to\infty} f^N(\omega) = f^\star(\omega)$ exists for all $\omega \in \Omega'$. Consider $\omega, \omega' \in \Omega'$. Since $h_n^\star(\omega) = h_n^\star(\omega')$, we have the following for arbitrary n, N, M:

$$\begin{aligned}
|\, f^\star(\omega) - f^\star(\omega')\,| &\leq |\, f^\star(\omega) - f^N(\omega)\,| + |\, f^N(\omega) - h_n^N(\omega)\,| \\
&\quad + |\, h_n^N(\omega) - h_n^\star(\omega)\,| + |\, h_n^\star(\omega') - h_n^M(\omega')\,| \\
&\quad + |\, h_n^M(\omega') - f^M(\omega')\,| + |\, f^M(\omega') - f^\star(\omega')\,|
\end{aligned}$$

Fix an arbitrary $\epsilon > 0$. Choose n so that $\|f - h_n\| < \frac{\epsilon}{6}$. We can argue as in the proof of the Lemma A.15 to show that this implies $|\, f^N(\omega) - h_n^N(\omega)\,| < \frac{\epsilon}{6}$ and $|\, f^M(\omega') - h_n^M(\omega')\,| < \frac{\epsilon}{6}$ for all N, M. Since $f^N(\omega) \to f^\star(\omega)$ and

$h_n^N(\omega) \to h_n^\star(\omega)$, we can choose N so that $|\, f^\star(\omega) - f^N(\omega)\,| < \frac{\epsilon}{6}$ and $|\, h_n^N(\omega) - h_n^\star(\omega)\,| < \frac{\epsilon}{6}$. Similarly, we can choose M such that $|\, f^\star(\omega') - f^M(\omega')\,| < \frac{\epsilon}{6}$ and $|\, h_n^M(\omega') - h_n^\star(\omega')\,| < \frac{\epsilon}{6}$. Therefore, $|\, f^\star(\omega) - f^\star(\omega')\,| < \epsilon$. Since ϵ was arbitrary, $f^\star(\omega) = f^\star(\omega')$. ■

For each $A \subset S$, define $I_A(\omega) = 1$ if $s_1 \in A$ and 0 otherwise.

Lemma A.17 *If $\succcurlyeq$ is ergodic, there exists Ω' such that for all $\omega \in \Omega'$ and all $A \in \mathcal{S}^\circ$, the limit $I_A^\star(\omega)$ exists and it is constant in Ω'.*

Proof: It is clear that $I_A \in B(\Omega, \mathcal{S}^1)$ for every $A \in \mathcal{S}^\circ$. By Lemmas A.14 and A.16, there is Ω' such that the desired properties hold for all $A \in \mathcal{S}^\circ$. ■

From now on, let the set Ω' given by the lemma above be fixed.

Lemma A.18 *There exists a finitely additive and monotone set function $\nu : \mathcal{S}^\circ \to [0,1]$ such that $I_A^\star(\omega) = \nu(A)$ for every $\omega \in \Omega'$ and $A \in \mathcal{S}^\circ$. Moreover, $I_A \sim I_A^\star \sim \nu(A)$.*

Proof: If $A_1, ...A_n$ are disjoint events in $\mathcal{S}^\circ$, then $I^\star_{\cup_{i=1}^n A_i}(\omega)$ is equal to:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} I_{\cup_{i=1}^n A_i}(T^j\omega) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} \sum_{i=1}^{n} I_{A_i}(T^j\omega) = \sum_{i=1}^{n} I^\star_{A_i}(\omega), \quad (17)$$

for all $\omega \in \Omega'$. This shows that the limit $I_A^\star(\omega)$ exists and it is unique for all $\omega \in \Omega'$ and all $A = \cup_{i=1}^n A_i$, for $A_1, ...A_n$ disjoint events in $\mathcal{S}^\circ$, that is, for all $A \in \mathcal{S}^\circ$. For each $A \in \mathcal{S}^\circ$, define $\nu(A) \equiv I_A^\star(\omega)$ (for any $\omega \in \Omega'$). If $A, B \in \mathcal{S}^\circ$ and $A \subset B$, then $I_A^\star(\omega) \leq I_B^\star(\omega)$, that is, ν is monotone. (17) shows that ν is finitely additive. Finally, $I_A \sim I_A^\star$ from the Subjective Ergodic Theorem and $I_A^\star \sim \nu(A)$ because $I_A^\star = \nu(A)$ off a null set $(\Omega \setminus \Omega')$. ■

Let ν defined in the proof of the above lemma be fixed. We have already shown that ν is finitely additive, but more is true:

Lemma A.19 *$\nu : \mathcal{S}^\circ \to [0,1]$ is countably additive.*

Proof: Consider a decreasing sequence of sets $A_n \in \mathcal{S}^\circ$, $A_n \downarrow \emptyset$. By Billingsley (1995), Example 2.10, p. 25, it is sufficient to prove that $\nu(A_n) \to \nu(\emptyset) = 0$. Suppose otherwise. Then there exist $\epsilon > 0$ and a subsequence A_{n_j} such that $\nu(A_{n_j}) \geq \epsilon$, which means that $I^\star_{A_{n_j}} \succcurlyeq \epsilon$. It is clear that $I^\star_{A_{n_j}}$ converges to 0 pointwise and, therefore, $I^\star_{A_{n_j}} \to 0$. Assumption 3 implies that $0 \succcurlyeq \epsilon$, but this contradicts Assumption 5. ■

Lemma A.20 *There exists a unique extension of ν to $\mathcal{S}$.*

Proof: By the Caratheodory extension theorem (see Royden (1968, Theorem 8, p. 257)), the following outer measure is the unique extension of ν to all $A \in \mathcal{S}$:

$$\nu(A) \equiv \inf\left\{\sum_{n=1}^{\infty} \nu(A_n) : A_n \in \mathcal{S}^\circ, A \subset \cup_{n=1}^{\infty} A_n\right\}. \tag{18}$$

■

Lemma A.21 *For any $A \in \mathcal{S}$, $I^\star_A(\omega) = \nu(A)$ for all $\omega \in \Omega'$. Therefore, $I^\star_A \sim \nu(A)$ for all $A \in \mathcal{S}$.*

Proof: Take $A_i \in \mathcal{S}^\circ$, such that $A \subset \cup_{i=1}^{n} A_i$. It is clear that $I_A(\omega) \leq \sum_{i=1}^{n} I_{A_i}(\omega)$ and, therefore,

$$\bar{I}_A(\omega) \equiv \limsup_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} I_A(T^j\omega) \leq \sum_{i=1}^{n} I^\star_{A_i}(\omega) = \sum_{i=1}^{n} \nu(A_i).$$

Since this is valid for any n, (18) implies that $\bar{I}_A(\omega) \leq \nu(A)$. Similarly,

$$\underline{I}_A(\omega) \equiv \liminf_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} I_A(T^j\omega) \leq \sum_{i=1}^{n} I^\star_{A_i}(\omega) = \sum_{i=1}^{n} \nu(A_i),$$

which proves that $\underline{I}_A(\omega) \leq \nu(A)$. It is easy to see that $\bar{I}_A(\omega) = 1 - \underline{I}_{A^c}(\omega)$. Assume that for some A and $\omega \in \Omega'$, we have $\bar{I}_A(\omega) < \nu(A)$. Then, $\underline{I}_{A^c}(\omega) = 1 - \bar{I}_A(\omega) > 1 - \nu(A) = \nu(A^c)$, but this contradicts $\underline{I}_{A^c}(\omega) \leq \nu(A^c)$. This shows that $\bar{I}_A(\omega) = \nu(A)$. Similarly, $\underline{I}_A(\omega) = \nu(A)$, for all $\omega \in \Omega'$, which shows that the limit $I^\star_A(\omega)$ exists and it is equal to $\nu(A)$ for all $\omega \in \Omega'$. The conclusion $I^\star_A \sim \nu(A)$ for all $A \in \mathcal{S}$ follows trivially. ■

Let $\mathcal{F}_1$ denote the set of acts measurable with respect to the firs coordinate.

Lemma A.22 *The preference $\succcurlyeq$ restricted to $\mathcal{F}_1$ is an expected utility preference with subjective probability ν.*

Proof: Let $A_1, ..., A_n \in \mathcal{S}$. By Lemma A.21, $\sum_{i=1}^n \alpha_i I^\star_{A_i}(\omega) = \sum_{i=1}^n \alpha_i \nu(A_i)$ for all $\omega \in \Omega'$. Therefore, $\sum_{i=1}^n \alpha_i I_{A_i} \sim \sum_{i=1}^n \alpha_i I^\star_{A_i} \sim \sum_{i=1}^n \alpha_i \nu(A_i)$, that is, $f \sim \int f d\nu$ for every simple function. Now, given $f \in \mathcal{F}_1$, let $B > 0$ be a bound for f, that is, $f(\omega) \in [-B, B]$, $\forall \omega \in \Omega$. For each $j \in \{-2^n, -2^n + 1, ..., 2^n\}$ and $n \in \mathbb{N}$, let $A_{j,n} \equiv \{\omega : f(\omega) \in [\frac{jB}{2^n}, \frac{(j+1)B}{2^n})\}$. Define:

$$f^n(\omega) = \sum_{j=-2^n}^{2^n} \frac{jB}{2^n} 1_{A_{j,n}}(\omega),$$

that is, $f^n(\omega)$ is valued $\frac{jB}{2^n}$, whenever $\frac{jB}{2^n} \leqslant f(\omega) < \frac{(j+1)B}{2^n}$. It is easy to see that f^n is a sequence of simple functions that converge pointwise to f, $f^n \uparrow f$. Also, by the Lebesgue Monotone Convergence Theorem, $\int f^n d\nu \to \int f d\nu$. Since $\{\omega : \lim_{n\to\infty} f^n(\omega) \neq f(\omega)\} = \emptyset$ and the empty set is $\succcurlyeq$-null because $\succcurlyeq$ is reflexive, then $f^n \to f$. Since $f^n \sim \int f^n d\nu$, we conclude by continuity that $f \sim f^\star \sim \int f d\nu$. Then, by Assumptions 2 and 5, we have $f \succcurlyeq g \Leftrightarrow \int f d\nu \geq \int g d\nu$, that is, the preference is an Expected Utility. ■

The following lemma completes the proof of the theorem:

Lemma A.23 *The preference $\succcurlyeq$ is an expected utility preference with subjective probability equal to the the product measure ν^∞ of ν.*

Proof: Since Ω is also Polish, Lemma A.13 guarantees the existence of a countable algebra Σ° that generates Σ. Let $B(\Omega, \Sigma^\circ)$ denote the space of all uniform limits of finite linear combinations of characteristic functions of sets in Σ°. Now, it is sufficient to adapt the arguments of Lemmas A.14 through A.22 to conclude that the preference is expected utility with a probability μ. If $A \subset \Omega$ is invariant, the fact that $\succcurlyeq$ is ergodic implies that A is $\succcurlyeq$-trivial and (the adaptation of) Lemma A.21 implies that $\mu(A) = 0$ or $\mu(A) = 1$. This is equivalent to say that μ is ergodic. By Hewitt and Savage (1955), Theorem 5.3, p. 478, μ is the i.i.d.product of some measure ν, that is, $\mu = \nu^\infty$, which concludes the proof. ■

A.5 Proof of Theorem 4

By Theorem 2, $f^\star(\omega) = \int f dP^{\theta(\omega)} = \Phi_\Theta(f)(\omega)$ for $P^{\theta(\omega)}$-almost all $\omega \in E^\theta$. Let C be the set of all $\omega \in \Omega'$ such that $f^\star(\omega) \neq \int f dP^{\theta(\omega)}$. We have just argued that $P^\theta(C) = 0$ for all $\theta \in \Theta$. By Assumption 6, C is null. Therefore, $f^\star \sim \Phi_\Theta(f)$. Since $f \sim f^\star$ by the Subjective Ergodic Theorem, transitivity concludes the proof.

A.6 Proof of Theorem 5

Proof of Theorem 5: Let $\mathcal{C}$ be the family of all subsets of S which are $\succcurlyeq$-statistically unambiguous. Note that S and $\emptyset$ are in $\mathcal{C}$, and since S is finite, we can write $\mathcal{C} = \{A^0, A^1, ..., A^n\}$, with $A^0 = \emptyset$ and $A^n = S$. Now, we will show that $\mathcal{C}$ is a λ-system, by arguing that it is closed for complementation and disjoint unions.

Note that if $A \in \mathcal{C}$ then $A^c \in \mathcal{C}$. Indeed, if the limit $1^\star_{A_\Omega}(\omega)$ exists and it is equal to γ, then the limit $1^\star_{A^c_\Omega}(\omega)$ also exists and it is equal $1-\gamma$. If A and B are in $\mathcal{C}$, let Ω_A and Ω_B be the respective sets with $\succcurlyeq$-null complements such that the limits $\nu(A,\omega)$ and $\nu(B,\omega)$ respectively exist. Then, for all $\omega \in \Omega_{A\cup B} \equiv \Omega_A \cap \Omega_B$, the limit $\nu(A\cup B,\omega)$ exist and it is constant and equal to $\nu(A,\omega)+\nu(B,\omega)$. Therefore, $A \cup B \in \mathcal{C}$.[41]

For each $A^k \in \mathcal{C}$, consider the algebra $\mathcal{S}^k$ given by the sets $\mathcal{S}^k \equiv \{\emptyset, A^k, S\setminus A^k, S\}$. Notice that the ergodicity of the preference was used in the proof of Theorem 3 only in the first paragraph of the proof of Lemma A.16, to argue that the limit is constant in a subset of Ω with $\succcurlyeq$-null complement. This is true for all sets in $\mathcal{S}^k$, since they are $\succcurlyeq$-statistically unambiguous. Therefore, we can repeating the proof of Theorem 3 for each $\mathcal{S}^k$ (instead of $\mathcal{S}$), and obtain the existence of $\Omega^k \subset \Omega$ and a measure ν_k defined in $\mathcal{S}^k$ for $k = 1, ..., n$ such that $\nu_k(A) = 1^\star_{A_\Omega}(\omega)$, for all $A \in \mathcal{S}^k$ and $\omega \in \Omega^k$. Put $\Omega' \equiv \cap_{k=1}^n \Omega^k$. It is clear that Ω' has a $\succcurlyeq$-null complement. If $A \in \mathcal{S}^i \cap \mathcal{S}^j$, then $\nu_i(A) = \nu_j(A) = 1^\star_{A_\Omega}$, by construction. Thus, we can define a partial measure ν on $\mathcal{C} = \cup_{i=1}^n \mathcal{S}^i$ by putting $\nu(A) = \nu_k(A)$ if $A \in \mathcal{S}^k$. This measure is clearly additive, that is, $A, B \in \mathcal{C}$, with $A \cap B = \emptyset$ implies $\nu(A \cup B) =$

[41] Note that we cannot reproduce the same argument for intersections.

$\nu(A)+\nu(B)$. The representation for each f, g which are $\mathcal{C}$-measurable follows from an adaptation of the proof of Lemma A.22.

It remains to prove only that ν is actually a partial probability on $\mathcal{C}$, that is, there exists a extension of ν to $\mathcal{S}$. We have already argued that ν is additive in $\mathcal{C}$. Therefore, it is a real partial charge (Definition 3.2.1, p. 64 of Bhaskara Rao and Bhaskara Rao (1983)) and by Theorem 3.2.5, p. 65 of Bhaskara Rao and Bhaskara Rao (1983), it admits an extension to $\mathcal{S}$. This concludes the proof. ∎

To illustrate the theorem, consider some preferences considered in the literature, such as Choquet Expected Utility (CEU), Maxmin Expected Utility (MEU) or Bewley's. In CEU with capacity υ, define, following de Castro and Chateauneuf (2008) (see also Nehring (1999))

$$\mathcal{S}' \equiv \{A \in \mathcal{S} : \upsilon\left(B\right) = \upsilon\left(B \cap A\right) + \upsilon\left(B \cap A^c\right), \forall B \in \mathcal{S}\}.$$

This class of events is a σ-algebra and it may be called the class of the unambiguous events. de Castro and Chateauneuf (2008) shows that a CEU preference restricted to the events in $\mathcal{S}'$ is just a EU.

In MEU and Bewley's, there is a compact convex set $\mathcal{P}$ of probabilities that define the preference. For this set, define the following:

$$\mathcal{S}' \equiv \{A \in \mathcal{S} : \pi(A) = \pi'(A), \forall \pi, \pi' \in \mathcal{P}\}.$$

Again, $\mathcal{S}'$ is called the class of unambiguous events.[42] It is not difficult to see that the MEU and Bewley preferences are just expected utility preferences in $\mathcal{S}'$. This shows that (restricted) ergodicity is relevant even if the preferences allow for ambiguity, as those cited.

A.7 Remaining Proofs

Corollary 3 (Characterization of Exchangeability) *For every preference $\succcurlyeq$ satisfying Assumption 6 for the finitely based acts, the following are equivalent:*

[42] Although in the CEU case, $\mathcal{S}'$ is a σ-algebra, in MEU or Bewley's $\mathcal{S}'$ may be just a λ-system. A class $\mathcal{L}$ of subsets of X is a λ-system if the following conditions are satisfied: (i) $X \in \mathcal{L}$; (ii) $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$; (iii) $A_1, A_2, ... \in \mathcal{L}$ and $A_n \cap A_m = \emptyset$, imply $\cup_n A_n \in \mathcal{L}$. See de Castro and Chateauneuf (2008) for a discussion.

- *$\succcurlyeq$ is exchangeable.*

- *For every $f \in \mathcal{F}$, $f^\star$ is well-defined, except on a $\succcurlyeq$-null event, and $f \sim f^\star$.*

Proof of Corollary 3: The first implication comes from the Subjective Ergodic Theorem. Now assume that $f^\star$ is well-defined except in a $\succcurlyeq$-null set, and that $f \sim f^\star$. We want to prove that $\succcurlyeq$ is exchangeable. Fix a $f \in \mathcal{F}$. By assumption, $\left(\frac{1}{n}\sum_{i=1}^{n} f \circ \pi^i(\cdot)\right) \sim \left(\frac{1}{n}\sum_{i=1}^{n} f \circ \pi^i(\cdot)\right)^\star$.

Let $\left\{\left(E^\theta, P^\theta\right)\right\}_{\theta\in\Theta}$ be the parametrization given by Theorem 2. Since $P^\theta = P^\theta \circ \pi$ for all permutation $\pi \in \Pi$, we have, for P^θ-almost all $\omega \in E^\theta$:

$$f^\star(\omega) = \int_{E^\theta} f\, dP^\theta = \int_{E^\theta} f\, d(P^\theta \circ \pi^{-1}) = \int_{E^\theta} f \circ \pi\, dP^\theta = (f \circ \pi)^\star(\omega).$$

Let C be the set of all $\omega \in \Omega'$ such that $f^\star(\omega) \neq (f \circ \pi^i)^\star$ for some $i = 1, ..., n$. We have just argued that $P^\theta(C) = 0$ for all $\theta \in \Theta$. By Assumption 6, C is null. For all $\omega \notin C$,

$$\left(\frac{1}{n}\sum_{i=1}^{n} f \circ \pi^i(\omega)\right)^\star = \frac{1}{n}\sum_{i=1}^{n}\left(f \circ \pi^i\right)^\star(\omega) = \frac{1}{n}\sum_{i=1}^{n} f^\star(\omega) = f^\star(\omega).$$

Since C is null, this shows that

$$\left(\sum_{i=1}^{n} \alpha_i f \circ \pi^i(\cdot)\right) \sim f^\star \sim f,$$

as we wanted to show. ∎

The next lemma shows that under expected utility, continuity is equivalent to countable additivity:

Lemma A.24 *Assume that the preference is an expected utility given by a finitely additive probability μ, that is, $f \succcurlyeq g \Leftrightarrow \int f\, d\mu \geq \int g\, d\mu$. Then, Assumption 3 holds if and only if μ is countably additive.*

Proof: Let Assumption 3 hold and let $f^n = 1_{A_n}$, where A_n is a decreasing sequence of sets, with $A_n \downarrow \emptyset$. We want to prove that $\int f^n\, d\mu = \mu(A_n) \to$

$\int 1_\emptyset \, d\mu = \mu(\emptyset) = 0$, because this convergence implies that μ is countably additive (see for instance Billingsley (1995),Example 2.10, p. 25). Suppose otherwise. Then, there exists $\epsilon > 0$ and a subsequence f^{n_j} such that $\int f^{n_j} \, d\mu \geq \epsilon$, which means that $f^{n_j} \succcurlyeq \epsilon$. It is clear that f^{n_j} converges to $f = 1_\emptyset = 0$ pointwise and, therefore, $f^{n_j} \to f$. Assumption 3 implies that $0 \succcurlyeq \epsilon$, which is an absurd.

For the converse let sequences f^n and g^n satisfy: (i) $f^n \to f$ and $g^n \to g$; (ii) $|f^n(\omega)| \leq b(\omega)$ and $|g^n(\omega)| \leq b(\omega)$, for all ω and some $b \in \mathcal{F}$; (iii) $f^n \succcurlyeq g^n$, that is, $\int f^n \, d\mu \geq \int g^n \, d\mu$. Then the assumptions of the Lebesgue Convergence Theorem are satisfied and this implies $\int f^n \, d\mu \to \int f \, d\mu$ and $\int g^n \, d\mu \to \int g \, d\mu$. Therefore, $\int f \, d\mu \geq \int g \, d\mu$, that is, $f \succcurlyeq g$. ■

References

Al-Najjar, N. I. (2009): "Decision Makers as Statisticians: Diversity, Ambiguity and Learning," *Econometrica*, forthcoming.

Bewley, T. (1986): "Knightian Decision Theory: Part I," Cowles Foundation Discussion Paper no. 807.

Bewley, T. (2002): "Knightian decision theory. Part I," *Decisions in Economics and Finance*, 25(2), 79–110.

Bhaskara Rao, K. P. S., and M. Bhaskara Rao (1983): *Theory of Charges*. Academic Press Inc., New York.

Billingsley, P. (1965): *Ergodic theory and information*. Wiley.

Billingsley, P. (1995): *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edn., A Wiley-Interscience Publication.

de Castro, L. I., and A. Chateauneuf (2008): "Ambiguity Aversion and Trade," Discussion paper, University of Illinois at Urbana-Champaign.

de Finetti, B. (1937): "La prévision: ses lois logiques, ses sources subjectives," in *Annales de l'Institut Henri Poincaré*, vol. 7, pp. 1–68.

——— (1938): "Sur la condition d'equivalence partielle," *Translated in Studies in Inductive Logic and Probability II*, 739.

——— (1989): "Probabilism," *Erkenntnis*, 31(2), 169–223, (Originally published in 1931 as *"Probabilismo"* in Italian).

Diaconis, P. (1992): "Sufficiency as Statistical Symmetry," in *Mathematics Into the Twenty-First Century: 1988 Centennial Symposium, August 8-12*. American Mathematical Society.

Diaconis, P., and D. Freedman (1980): "de Finetti's theorem for Markov chains," *Ann. Probab*, 8(1), 115–130.

——— (1984): "Partial exchangeability and sufficiency," *Proceedings of the Indian Statistical Intitute Golden Jubilee International conference on Statistics: Applications and New Directions*, pp. 205–236.

Dunford, N., and J. Schwartz (1958): *Linear Operators, Part I.* Interscience, New York.

Epstein, L., and K. Seo (June 2008): "Symmetry of Evidence without Evidence of Symmetry," Boston University.

Ghirardato, P., F. Maccheroni, M. Marinacci, and M. Siniscalchi (2003): "A Subjective Spin on Roulette Wheels," *Econometrica*, 71(6), 1897–1908.

Gilboa, I., F. Maccheroni, M. Marinacci, and D. Schmeidler (2008): "Objective and Subjective Rationality in a Multiple Prior Model," Collegio Carlo Alberto, Universita di Torino.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Nonunique Prior," *J. Math. Econom.*, 18(2), 141–153.

Gray, R., and L. Davisson (1974): "The ergodic decomposition of stationary discrete random processes," *Information Theory, IEEE Transactions on*, 20(5), 625–636.

Hewitt, E., and L. J. Savage (1955): "Symmetric measures on Cartesian products," *Trans. Amer. Math. Soc.*, 80, 470–501.

Hu, T. (2008): "Expected Utility Theory from the Frequentist Perspective," *Economic Theory*, Forthcoming.

Jackson, M. O., E. Kalai, and R. Smorodinsky (1999): "Bayesian Representation of Stochastic Processes under Learning: de Finetti Revisited," *Econometrica*, 67(4), 875–893.

Kreps, D. (1988): *Notes on the Theory of Choice.* Westview Press.

Lauritzen, S. (1984): "Extreme Point Models in Statistics," *Scandinavian Journal of Statistics*, pp. 65–91.

Löh, C. (2006): "The Ergodic Decomposition Theorem," WWU Münster.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Malevolent Nature, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–98.

Machina, M. J., and D. Schmeidler (1992): "A more robust definition of subjective probability," *Econometrica*, 60(4), 745–780.

Nehring, K. (1999): "Capacities and probabilistic beliefs: A precarious coexistence," *Mathematical Social Sciences*, 38(2), 197–214.

Royden, H. L. (1968): *Real Analysis*. MacMillan Publishing Co., Inc., New York, 2 edn.

Savage, L. J. (1954): *The foundations of statistics*. John Wiley & Sons Inc., New York.

Sims, C. (1996): "Macroeconomics and Methodology," *The Journal of Economic Perspectives*, 10(1), 105–120.

Varadarajan, V. (1963): "Groups of automorphisms of Borel spaces," *Trans. Amer. Math. Soc*, 109(2), 191–220.

Viana, M. (2008): "Disintegration into conditional measures: Roklin's theorem," available at http://w3.impa.br/%7Eviana/out/rokhlin.pdf.

Zhang (1999): "Qualitative Probabilities on λ-Systems," *Mathematical Social Sciences*, 38, 11–20.