STATES, MODELS AND INFORMATION: A RECONSIDERATION OF ELLSBERG'S PARADOX

MASSIMILIANO AMARANTE

ABSTRACT. We show that a general process of decision making involves uncertainty about two different sets: the domain of the acts and another set, which we call the set of models for the decision maker. We study the effect of different information structures on the set of models, and prove the existence of a dichotomy: either the decision maker's ranking of the acts obeys Subjective Expected Utility theory or there are many events to which probabilities cannot be assigned. We use this result to formalize the idea of Knightian Uncertainty. The relevance of information structures associated to Knightian Uncertainty is shown by means of examples, one of which is a version of Ellsberg's experiments. Our findings show that a decision maker faces, generally speaking, uncertainties of two different types – "uncertainty about which state obtains" and "uncertainty about how the world works" – and that Savage's theory considers only uncertainty of the first type. Finally, in situations of Knightian Uncertainty, we identify the class of events to which probabilities can be assigned, and study the relation with the class of unambiguous events in the sense of [13] and [25].

1. INTRODUCTION

Several branches of economic theory (GEE, Game Theory, etc.) have developed relying, for the most part, on the assumption that individuals make decisions in accordance to Savage's theory of decision making under uncertainty [28]. Nevertheless, new theories have emerged over the past twenty five years: Prospect Theory [19], Choquet Expected Utility [29], Bewley's theory [4], Maxmin Expected Utility [14] and various generalizations of the latter [13]. All these belong to the class of Multiple Prior Models, and have encountered the favor of many decision theorists. The reason is clear: these theories try to deal with phenomena which are important and which cannot be dealt with by Savage's theory. The best known of these phenomena is that of Knightian Uncertainty. Summarily, it refers to situations where the information available to the decision maker is so coarse that it generates an inherent inability to assign probabilities to many events. This has been the main idea behind most Multiple Prior Models: " ... The idea is simple and appealing. Since the decision maker does not have enough information to form a meaningful single prior, he uses a set of priors consisting of all those priors compatible with

---

Date: December 2005. E-mail: ma734@columbia.edu
I benefited from comments and suggestions from Paolo Ghirardato, Tzachi Gilboa, Fabio Maccheroni, Massimo Marinacci, Rich McLean, Ket Richter, Aldo Rustichini, Paolo Siconolfi, Max Stinchcombe, William Thomson and Bill Zame. I also wish to thank seminar participants at Yale University, University of Iowa, RUD 2003, CalTech, University of Minnesota, University of Texas-Austin, Columbia University, University of Rochester, Rutgers University, University of Turin, Ohio State University.

his limited information" (Marinacci [24]). A statement of this sort, while of value as an inspiring principle, is hard to substantiate in the context of existing theories. In fact, nothing in those theories tells us (1) What situations should be classified under the label "Knightian Uncertainty"; (2) Why it is that Multiple Prior Models should pop up in such situations; and (3) Whether or not we can attribute to those theories the same normative value that we attribute to Savage theory in risky situations. The aim of the present paper is to contribute toward answering these types of questions. Our contribution is twofold. First, we will give a complete answer to the first problem, and we will discuss how our findings could be extended in order to answer to the other two problems. We intend to pursue such an extension in a future paper. Second, we will show that the notion of *set of states* as introduced in Savage is too restrictive in that it cannot account for all the uncertainty faced by a decision maker. This has two relevant implications. On the one hand, it shows that some of the limitations of Savage theory emerge precisely as a consequence of such a restrictive view of uncertainty. On the other hand, it paves the way for a theory of decision making capable of accommodating a whole new set of phenomena.

The stress we have been placing on the problem of Knightian Uncertainty should not be misinterpreted as a statement that Savage theory's only limitation is its inability to deal with Knightian Uncertainty. In fact, many others limitations have emerged over the years: Unforeseen Contingencies (Dekel, Lipman and Rustichini [8]), Case-Based decision theory (Gilboa-Schmeidler [15], [16]), Conditional choice (Fishburn,[11]), etc. These issues are as fundamental as that of Knightian Uncertainty. Some of them can be meaningfully addressed in the framework emerging from this paper. However, since the main objective is the identification of situations of Knightian uncertainty, we will limit ourselves to summary hints.

1.1. **Outline of the paper.** The paper takes off from the simple observation that the concept of "Information" has no explicit place in existing theories. This is what makes it hard to identify situations of Knightian uncertainty, for instance. Hence, the necessity of explicitly introducing Information into the model. We do so as follows. We begin by observing that – as shown in [2] – virtually any model of decision making we know of is associated to a certain structure, and that this does allow us to talk about Information in a meaningful and interesting way. Next, we formally introduce Information into the model, and study the properties associated to different types of Information structures. In doing so, we identify certain properties that provide the formal counterpart to the intuitive idea of Knightian Uncertainty. The relevance of information structures associated to Knightian uncertainty is shown by means of examples, one of which is a version of Ellsberg's paradox.

The paper introduces a number of new concepts which often either require lengthy explanations or lend themselves to quite novel interpretations. While this is an important part of our work, we wanted to avoid that too many detours would make the reader lose sight of the target. Because of this, we have opted in favor of relegating most of this material into the final sections. The remainder of this section presents an overview of the main steps that are performed in the paper.

1.1.1. *Representation theorems.* The theory of decision making under uncertainty is concerned with individuals choosing among a set of available alternatives. The outcome associated to each choice depends on the realization of a state of the world, $s \in S$, which is unknown to the decision maker. A representation theorem is a statement of the form "if the decision maker's ranking of the alternatives, $\succsim$, obeys certain rules, then that ranking can be represented by means of a functional $I : B(S) \to \mathbb{R}$ having a certain form".[1]

$$\begin{array}{ccc} B(S) & & \\ & I \searrow & \\ & & \mathbb{R} \end{array}$$

Figure 1

Thus, the functional I represents the decision maker in the sense that we can think of him *as if* he used I to rank the alternatives.

In [2], it was shown that, in essentially any axiomatic model, the functional I can be thought of as consisting of two parts as in the diagram below.

$$\begin{array}{ccc} B(S) & \xrightarrow{\kappa} & B(M) \\ & I \searrow & \downarrow V \\ & & \mathbb{R} \end{array}$$

Figure 2

The set M and the mapping κ are essentially the same in any model of decision making, while the functional V varies with the axioms one imposes on the decision maker's preference relation.

What does this buy us? Recognizing that the process of decision making can always be described as in the previous diagram generates advantages on two fronts. First, it allows us to provide new interpretations for existing theories (see [2], where the process of decision making is interpreted as a form of analogical reasoning). Second, it shows that the structure involved in the process of decision making is richer than previously thought (compare Fig. 1 to Fig. 2). This paves the way for a reconsideration of both the meaning of the term "Uncertainty" and of the role played by "Information" in decision problems.

1.1.2. *Two types of information.* Figure 2 states that the process of decision making involves two sets: the domain of the acts S and another set M. In [2], it was shown that M can always be represented as a set of probabilistic descriptions of the domain of the acts. Thus, the decision maker's ranking emerges as the outcome of two types of assessments: assessments about the likelihood of the various states and about the likelihood of the various probabilistic descriptions. Having realized that a decision maker can be uncertain about two different objects leads to the obvious observation that he can obtain valuable information about any of those sets. Hence, one is led to inquire about how information about M affects the decision maker's ranking of the alternatives. We remark that Ellsberg's experiments – where the decision maker is given information about the configuration of the urn rather than the set of states – are stripped down examples bringing to light that information about the set M may affect the decision maker's choices.

---

[1] Here $B(S)$ denotes the set of bounded (measurable) real-valued functions on S.

1.1.3. *Coarse information about M.* If we assume that the functional V in Fig. 2 is linear, then Theorem 1 of Section 3 delivers immediately the following result: *if the information about M available to the decision maker is sufficiently fine*[2], *then the decision maker ranks acts according to a Subjective Expected Utility criterion.* Yet, this does not say anything about our main concern, which is about what happens in those situations characterized by the presence of coarse information (cf. Marinacci's statement above). The study of such situations is the main theme beginning from Section 6 on. Here is the basic idea: we take as a benchmark a situation in which the information about M is fine and the Subjective Expected Utility (SEU) theorem obtains; then, we vary the information about M, and study whether or not our SEU theorem still obtains. More precisely, given a σ-algebra $\mathcal{M}$ on M for which the SEU theorem obtains, we describe coarser information structures by means of partitions of M into measurable events. Each partition $\mathcal{I}$ expresses the restriction that, on the basis of his information, the decision maker has to make the same decision in correspondence of points of M which lie in the same cell of the partition. Equivalently, the decision maker's information is now described by the sub σ-field generated by the partition, and the decision process occurs according to the diagram

$$\begin{array}{ccc} B(S) & \longrightarrow & B(M/\mathcal{I}) \\ & I \searrow & \downarrow V \\ & & \mathbb{R} \end{array}$$

where $M/\mathcal{I}$ denotes the quotient space endowed with its canonical measurable structure. Our main result is the existence of a dichotomy: Either the partition $\mathcal{I}$ is such that the SEU theorem obtains or there are many events (and many acts) to which the decision maker cannot assign probabilities.

1.1.4. *Knightian uncertainty.* The idea of Knightian uncertainty is associated to the presence of a large degree of fuzziness. The decision maker does know that a number of things can happen, but he is unable to assess to which degree they are likely to happen. In even looser terms, it feels like any turn might lead anywhere and for no precise reason. We are going to see that all partitions for which the SEU theorem fails are associated to the very same property: they are nonmeasurable in the sense of Rokhlin [27] (see Definition 2, Section 7). This mathematical property is the precise translation of the idea of fuzziness: intuitively, it states that the decision maker, while aware that two things are different, is unable *on the basis of his information* to distinguish between the two. And, of course, if one cannot distinguish between two things, then there is no way to assign probabilities to them.

Formalizing the idea of Knightian uncertainty does not serve to the mere need of making a vague idea precise. Rather, it paves the way for a thorough study of both the idea itself and of those phenomena that are intuitively associated with it. For instance, now that we know that Knightian uncertainty corresponds to nonmeasurable information structures, we have a formal setting where questions like "Is there any reason why, in the presence of Knightian uncertainty, a rational decision maker should not conform to a SEU criterion" can be meaningfully addressed. As we shall say in the next paragraph, the formalization of the concept brings about

[2] In a sense made precise later.

some other advantages in that it casts a new light on parts of the theory that, at the outset, were probably not immediately related to it.

1.1.5. *Two types of uncertainty.* In Sections 9 and 10, we give two examples of nonmeasurable information structures. The first describes uncertainty about the class of measure zero events, the second is a version of Ellsberg's three-color urn experiment. The finding that these information structures produce (in our formal sense) Knightian uncertainty leads to a reinterpretation of Theorem 1 of Section 3. In light of those findings, the presence of the set M appears neither as an accident nor as a convenient devise. Quite to the contrary, the set M witnesses the presence of two distinct types of uncertainty faced by decision makers. The first is the classical one: uncertainty as to which state obtains. The second is uncertainty about the relations existing across the various events. Intuitively, this is uncertainty about how the world works, and we will show that this type of uncertainty has been neglected in Savage's formulation (equivalently, that Savage's notion of state is not as comprehensive as previously thought).

This completes the description of the paper's core. A few sections (12 to 15) conclude the paper. Mainly, they are devoted to either highlighting implications of our findings for other parts of the theory or to deepening and clarifying interpretations given before. Probably, Section 13 deserves special mention as we identify those events to which probabilities can be assigned. We characterize the class of these events – which we call *subjectively measurable* – and study the relation of our notion with the notion of *ambiguous event* proposed by [13] and [25].

2. Notation

The notation employed in the paper is standard. Alternatives available to the decision maker are viewed as mappings $S \longrightarrow X$, and are called acts. The domain of the acts S (also referred to as the set of states) is equipped with a σ-field Σ, and every act is Σ-measurable. The set X is the prize space, and is assumed to be a mixture space. In the next section, we will make assumptions guaranteeing the existence of a utility function $u : X \longrightarrow \mathbb{R}$. This produces an embedding of the acts into the set $B(S, \Sigma)$ of bounded real-valued Σ-measurable functions on S. In general, for $(C, \mathcal{C})$ an arbitrary measurable space, the notation $B(C, \mathcal{C})$ stands for the set of all bounded real-valued $\mathcal{C}$-measurable mappings on $(C, \mathcal{C})$. Sometimes, we just write $B(C)$ in the place of $B(C, \mathcal{C})$. The notation $ba_1(\Sigma)$ stands for the set of all finitely additive probability measures on Σ. The subset of $ba_1(\Sigma)$ consisting of all those probability measures which are countably additive is denoted by $\mathcal{P}(S)$.

Before abandoning this section, a couple of observations about the objects involved in this description are probably in order. Usually, we require that our theories be testable, at least in principle. This demands that the domain of the acts S be objectively given at least for an outside observer. Throughout the paper, we will stick with this interpretation. A similar observation can be made about the prize space X (see also [20]). The σ-field Σ must be given an objective meaning too, and cannot be interpreted as reflecting the decision maker's information about S. The latter, if anything, should be derived by the theory on the basis of the decision maker's behavior. One way to ensure that Σ has an objective meaning is as follows. Once acts have been identified to bounded real-valued functions, define

Σ as the coarsest σ-field which makes all the acts measurable. This interpretation will be maintained throughout the paper.

3. States and models

Let $\mathcal{F}$ denote the set of all acts, and $\mathcal{F}_c$ that of constant acts. Let $\succsim_S$ denote the decision maker's preference relation over $\mathcal{F}$. Under the following assumptions on $\succsim_S$

A1 $\succsim_S$ is complete and transitive.

A2 (C-independence) For all $f, g \in \mathcal{F}$ and $h \in \mathcal{F}_c$ and for all $\alpha \in (0,1)$

$$f \succ_S g \qquad \Longleftrightarrow \qquad \alpha f + (1-\alpha)h \succ_S \alpha g + (1-\alpha)h$$

A3 (Archimedean property) For all $f, g, h \in \mathcal{F}$, if $f \succ_S g$ and $g \succ_S h$ then $\exists \alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ_S g$ and $g \succ_S \beta f + (1-\beta)h$.

A4 (Monotonicity) For all $f, g \in \mathcal{F}$, $f(s) \succsim_S g(s)$ for any $s \in S \Longrightarrow f \succsim_S g$.

A5 (Non-degeneracy) $\exists x, y \in X$ such that $x \succ y$.

Ghirardato, Maccheroni and Marinacci (GMM, [13]) have shown

Theorem (GMM, [13]**)** $\exists$ a unique[3] weak*-closed set of measures $\mathcal{C} \subseteq ba_1(\Sigma)$ and a sup-norm continuous functional $I : B(S) \longrightarrow \mathbb{R}$ such that[4]

$$f \succsim_S g \qquad iff \qquad I(f) \geq I(g)$$

where

$$I(f) = \alpha(f) \min_{\nu \in \mathcal{C}} \int f d\nu + [1-\alpha(f)] \max_{\nu \in \mathcal{C}} \int f d\nu$$

We remark that the weak*-topology appearing in the Theorem is the one given by the duality $\sigma(ba_1(\Sigma), B(S, \Sigma))$. This result can be reformulated as follows

Theorem 1. *Given a preference relation satisfying Axioms 1 to 5 there exist*
(i) a measurable space $(M, \mathcal{M})$
(ii) an linear mapping $\kappa : B(S) \to B(M)$
(iii) a functional $V : B(M) \to \mathbb{R}$
such that

$$f \succcurlyeq_S g \qquad iff \qquad V \circ \kappa(f) \geq V \circ \kappa(g)$$

Moreover, one can take $M = \mathcal{C} \subset ba_1(\Sigma)$ *and* κ *is defined by* $f \longmapsto \phi_f$, *where* ϕ_f *is the function that at point* $\mu \in \mathcal{C}$ *takes the value*

$$\phi_f(\mu) = \int_S f d\mu$$

Finally, κ *is sup-norm to sup-norm continuous and* V *is sup-norm continuous.*

The first part of Theorem 1 was already proven in [2]. Here, we have added the continuity properties of κ and V, which will be needed later.

Theorem 1 states that we can think of the decision process as consisting of two stages: first, the decision maker maps the bets $B(S)$ into $B(M)$ by means of κ, then he orders $B(M)$, and hence $B(S)$, by means of V (see Fig. 2, Section 1). In other words, we can always think of the decision process as consisting of the two

---

[3] In the sense explained there.

[4] Axioms 1 to 5 guarantee the existence of a utility $u : X \longrightarrow \mathbb{R}$. For notational simplicity, in the statement as well as in the rest of the paper, we identify an act $f : S \longrightarrow X$ with the real-valued function $u \circ f$.

stages described in Fig. 2 above. We remark that theorems of this type obtain for a much wider class of preferences (see [2]).

3.1. **Comments.**

3.1.1. *M is a space of risky descriptions.* A notable feature of Theorem 1 (and of its generalizations, see [2]) is that the set M – which we call the *set of models for the decision maker* – can always be described by means of a set of probability measures on S. An additional insight comes from the observation that the mapping κ, which links S and M in the theorem, can always be defined in the same way. That allows us to think of each element in M as describing an ordering of the acts (on S) conforming to the Expected Utility criterion. To see this, just recall that if f is an act and m is a model, then f is evaluated by $\int f dm$ in correspondence to model m. In the spirit of the Knightian distinction between risk and ambiguity, we can then think of M as a collection of risky descriptions of the problem faced by the decision maker. Finally, we observe that M is not just a set, but rather a space in that it comes equipped with a measurable structure. This expresses the understanding that the decision maker has of M.

3.1.2. *Two types of information.* No matter how one interprets M, it is a set that is linked to S by a certain relation and the decision maker takes this into account when it comes to ordering the bets in $B(S)$. Then, Theorem 1 allows us to think of a decision maker who is uncertain about both S and M. The most useful implication of Theorem 1 stems from this simple observation. In fact, since the decision maker is uncertain about two objects, he can get information about any of them. In particular, if he gets some information about M, then this could affect his ranking of the bets $B(S)$, because M and S are linked to each other. In Sections 9 to 11, we will show that the simultaneous presence of S and M witness the simultaneous presence of uncertainties of inherently different types.

3.2. **Standardness assumptions.** We now introduce two additional assumptions which we maintain throughout the paper, unless otherwise stated.

A6 (Standard Setting) The set S has the cardinality of the continuum. Moreover, the measurable space (S, Σ) is a standard Borel space (see Appendix A.1 for a definition).

Let $\succsim$ be a preference relation satisfying axioms 1 to 5. Let $\succsim^*$ denote the unambiguous preference relation ([13], Sec. B.3) associated to $\succsim$.

A7 (Monotone Continuity, GMM [13]**)** For all $x, y, z \in X$ such that $y \succ^* z$, and all sequences of events $\{A_n\}_{n\geq 1} \subseteq \Sigma$ with $A_n \downarrow \emptyset$, there exists $\bar{n} \in \mathbb{N}$ such that $y \succ^* xA_{\bar{n}}z$.

Assumption 6 implies, by the Isomorphism Theorem for Borel spaces (see Appendix A.1) that, without any loss, the reader can think of (S, Σ) as of $[0, 1]$ equipped with the Lebesgue σ-algebra Λ. Axiom 7[5] is equivalent to the property that all the measures in Theorem 1 are countably additive ([13], Sec. B.3).

The main reason for introducing these assumptions is technical: they allow us to use the machinery of Polish group actions, which plays a big role in the next sections. We remark, however, that finite sets equipped with the maximal algebra are standard Borel spaces. We will study this case in Section 14.

[5]For a justification of axiom A7, the reader might consult [3].

3.3. **A Polish setting.** Assumption A6 and A7 allow us to prove a new version of Theorem 1 where the weak*-topology $\sigma(ba_1(\Sigma), B(\Sigma))$ is replaced by the weak*-topology $\sigma(\mathcal{P}(S), C_b(S))$, thus ensuring that M is a Polish space. This is the version that we are going to be using in the rest of the paper. We stress that from now on by weak*-topology on M, we mean the topology that M inherits as a subspace of $(\mathcal{P}(S), \sigma(\mathcal{P}(S), C_b(S)))$.

Theorem 2. *Let (S, Σ) be a standard space. Then, given a preference relation $\succsim_S$ satisfying Axioms 1 to 5 and Axiom 7 there exist*
(i) a Borel space $(M, \mathcal{M})$
(ii) a linear mapping $\kappa : B(S) \to B(M)$
(iii) a functional $V : B(M) \to \mathbb{R}$
such that

$$f \succcurlyeq_S g \qquad iff \qquad V \circ \kappa(f) \geq V \circ \kappa(g)$$

The set M is weak-compact in $(\mathcal{P}(S), C_b(S))$ and κ is defined by $f \longmapsto \phi_f$, where ϕ_f is the function that at point $m \in M$ takes the value*

$$\phi_f(m) = \int_S f dm$$

The following corollary to Theorem 2, while mainly of a technical nature, is of independent interest. It states that if an invariant biseparable preference (i.e., a preference satisfying A1 to A5, see [13]) satisfies the axiom of monotone continuity, then the set of priors is a Polish space (a separable, completely metrizable topological space).

Corollary 1. *Let (S, Σ) be a Standard Space. A preference relation $\succcurlyeq_S$ satisfies A1 to A5 and A7 if and only if there exists a sup-norm continuous functional $I : B(S) \longrightarrow \mathbb{R}$ such that*

$$f \succsim g \qquad iff \qquad I(f) \geq I(g)$$

where

$$I(f) = \alpha(f) \min_{\nu \in \mathcal{C}} \int f d\nu + [1 - \alpha(f)] \max_{\nu \in \mathcal{C}} \int f d\nu$$

Moreover, $\mathcal{C}$ is unique and it is a weak-closed subset of $(\mathcal{P}(S), \sigma(\mathcal{P}(S), C_b(S)))$.*

The result obtains by observing that the functionals V of Theorem 1 and V of Theorem 2 are defined (pointwise) in exactly the same way and that the continuity properties of the functional do not change by moving from Theorem 1 to Theorem 2.

4. An Illustration: Ellsberg's three-color urn experiment

Ellsberg's [10] three-color urn experiment is the perfect example to illustrate the concepts we have developed so far: the ingredients of the model emerging from Theorems 1 and 2 are all explicitly there. In the three-color urn experiment, a decision maker has to rank bets which pay a certain amount of money depending on the color of a ball which is drawn from an urn. He is told that the urn contains 90 balls, of which 30 are red (R) while the remaining are either blue (B) or green (G) in unknown proportion.

In our notation, the 90-ball urn is the set S, the domain of the bets. We can think of it as a set with 90 points. The set of bets is the set of nonnegative functions with the urn, S, as a domain. Σ, the coarsest σ-field which makes all the bets measurable, is the power set of S. The set of models M is the set of all possible configurations of the urn. Namely, the set of all possible combinations of R, B and G that add up to 90. Finally, $B(M)$ (considered as a set) can be viewed as a system of hypothetical bets on the configuration of the urn. In other words, these would be bets of the form, "I pay you \$$x$ if the number of blue balls is 46 and I give you 0 otherwise". Notice that these bets are different from the ones actually offered to the decision maker.

In the experiment, the decision maker is told that the only possible configurations are those where the number of balls that are either blue or green is 60. We stress that this is explicitly information about the set of models and not about the domain of the bets. It is clear, however, that the information is relevant for ranking the bets in $B(S)$. In particular, if the decision maker is going to use a probability measure on S, then he must make sure that the probability of the event $R \subset S$ is $1/3$. Besides these obvious considerations, it is not clear how information about the set of models translates into information about the domain of the bets. In fact, there is a piece missing in our exposition, the σ-field $\mathcal{M}$ on the set of models. This describes the decision maker's information which is that "the number of balls that are either blue or green is 60". Now, which σ-field on M describes this information? We will have to introduce a number of additional concepts before attacking this problem. We will do so in Section 10.

5. The Bayesian view: integrating over priors

The recognition that a decision maker might obtain information about M rather than about S is by no means sufficient to justify a departure from the Bayesian paradigm. In fact, a Bayesian would argue that uncertainty about M would be described by means of a probability measure, P, on that set; i.e., a measure over measures on Σ. Then, the decision maker would "average" these measures using the "weights" given by P and, by doing so, he would end up with a single measure on S, which describes his uncertainty. In terms of Fig. 2, this means that if you are Bayesian, then your functional V has to be linear. Since κ is always linear, $V \circ \kappa$ is automatically linear and SEU follows from Riesz representation theorem.

Everything seems to work fine, but ... let us take a closer look. When is it that this "Integration over Priors" business works in the way just described? Clearly, since we are talking about integration over M, we have been implicitly referring to an algebra defined on M. This is the one appearing in Theorem 2, namely the Borel σ-algebra generated by the weak*-topology. As it turns out (we shall see it a moment), this algebra is quite fine, which corresponds to the assumption that the decision maker's information about M is pretty good. But, the motivation for our work is exactly the statement that Multiple Prior Models would pop up when the decision maker's information is coarse! In this respect, the traditional Bayesian argument does not seem to provide a satisfactory answer: in fact, it does not seem to provide an answer at all.

6. What is information?

Generally speaking, information about a set X is a pair (X, x), where X is the set in question and x is a collection of subsets of X; x tells you that certain elements have a certain property, certain others do not, etc. It is crucial to distinguish between X as a set and (X, x) as a space, that is between the set and the information you are given about it. An example might clarify. Suppose that you are dealing with an individual who has never heard of the real line. In principle, you might give him the list of all real numbers. In particular, this contains the numbers e and π. Now, ask him to show you that e and π are actually different numbers. Since he has never heard of the real line, his information about it is described by the collection of subsets $\{\varnothing, \mathbb{R}\}$, and it is clear that, *on the basis of this information*, he has no way to show you that e and π are actually different. Of course, things would be different if he knew the natural topology of the line.

6.1. **Coarse information structures.** Let $\mathcal{M}$ denote the σ-algebra on M generated by the weak*-topology. Above, we saw that the "integration over priors" argument works if the decision maker's information about M is described by $(M, \mathcal{M})$. But, what happens if the information is coarser? The first necessary steps toward answering this question consist of expressing (a) the notion of coarser information; and (b) the behavioral constraints that appear in the presence of coarser information. Point (a) simply requires us to replace $\mathcal{M}$ with a proper sub-algebra, which leads to the following

Definition 1. *An information structure on M is a pair $(\mathcal{M}, \mathcal{I})$, where $\mathcal{M}$ is the σ-algebra on M generated by the weak*-topology and $\mathcal{I}$ is a partition of M.*

The partition (and the associated sub σ-field) states that the decision maker has only partial information about M. This corresponds to the following situation (see Billingsley [5], pp. 57-58 and pp. 427-29): on the basis of his information, the decision maker can construct a statistical experiment whose outcome would tell him (in a statistical sense) in which element of the partition the true model lies. However, such a decision maker would not be able, *on the basis of his information*, to construct an experiment capable of distinguishing among models lying in the same cell of the partition.

These limitations translate into limitations in the decision maker's ability of ordering functions in $B(M)$. These are of two types: first, the functional $V : B(M) \longrightarrow \mathbb{R}$ must respect the informational constraint described by the partition; second, the decision maker can only evaluate those functions in $B(M)$ for which "enough information" is available. Limitations of the first type take the following form. Let $\mathcal{I}$ be a partition of M, and denote by ι a generic element of the partition. Once the decision maker is informed that ι obtains, he must evaluate functions in $B(M)$ according to this information. For $\varphi \in B(M)$, let us denote by $V(\varphi \mid \iota)$ such an evaluation. The limitation that all that the decision maker can get to know is an element of $\mathcal{I}$ translates into the condition

$$\varphi, \psi \in B(M) \text{ and } V(\varphi \mid \iota) = V(\psi \mid \iota) \text{ for all } \iota \in \mathcal{I} \implies V(\varphi) = V(\psi) \tag{6.1}$$

That is, if two functions in $B(M)$ are evaluated in the same way in correspondence to each and every element of the partition, then they must be evaluated in the same way unconditionally. Condition (6.1) can be restated in a more useful way. Define a mapping $\tilde{\pi}_V : B(M) \longrightarrow \mathbb{R}^{M/\mathcal{I}}$ by $\varphi \longmapsto (V(\varphi \mid \iota))_{\iota \in \mathcal{I}}$; that is, each function $\varphi \in B(M)$ is associated to the real function defined on the quotient $M/\mathcal{I}$ that at point ι (viewed as point in the quotient) takes value $V(\varphi \mid \iota)$. Then, condition (6.1) says that the functional $V : B(M) \longrightarrow \mathbb{R}$ must be expressible by means of the diagram below[6]

$$\begin{array}{ccc} B(M) & \xrightarrow{\tilde{\pi}_V} & \mathbb{R}^{M/\mathcal{I}} \\ & V \quad \searrow & \downarrow V' \\ & & \mathbb{R} \end{array}$$

The second issue is whether or not the decision maker has enough information to rank all the functions in $\tilde{\pi}_V\,(B(M)) = \mathit{range}\;\tilde{\pi}_V$. This is a condition on the decision maker's σ-field on $M/\mathcal{I}$: by definition, functions that are not measurable with respect to the decision maker's σ-field on $M/\mathcal{I}$ cannot be evaluated. In particular, given that V has to respect the condition stated by the diagram above, if $\psi \in B(M)$ is such that $\tilde{\pi}_V(\psi)$ is not measurable, then ψ cannot be evaluated (i.e., ψ is not measurable with respect to the decision maker's information). Now, notice that the decision maker's σ-field on $M/\mathcal{I}$ must be such that the canonical projection $\pi : (M, \mathcal{M}) \longrightarrow M/\mathcal{I}$ is measurable. For, if not, we would reach the absurd conclusion that the decision maker actually has more information than the one described by $\mathcal{M}$. To see this, suppose, by the way of contradiction, that the decision maker has a field on $M/\mathcal{I}$ for which the canonical projection is not measurable. This means that there exists an event A in $M/\mathcal{I}$ for which $\pi^{-1}(A) \notin \mathcal{M}$. However, by knowing the set M and his field on $M/\mathcal{I}$, the decision maker has enough information to evaluate the bet $\chi_{\pi^{-1}(A)}$ ($\chi_{\cdot}$ denotes indicator functions); in fact, he would evaluate $\chi_{\pi^{-1}(A)}$ in the same way as he evaluates the bet χ_A. This is a contradiction because, by definition, the evaluation of bets like $\chi_{\pi^{-1}(A)}$ is not permissible on the basis of his information because $\pi^{-1}(A)$ is not an event.

Because of this, $M/\mathcal{I}$ is endowed with the *finest* σ-field which makes the canonical projection measurable, and we conclude that the decision maker can order the functions in $B(M)$ *while respecting his information* if and only if two conditions are satisfied:

(*) $\tilde{\pi}_V\,(B(M)) \subset B(M/\mathcal{I}, \mathcal{M}/\mathcal{I})$, where $\mathcal{M}/\mathcal{I}$ denotes the finest σ-field which makes the canonical projection measurable; and

(**) There exists $V' : B(M/\mathcal{I}) \longrightarrow \mathbb{R}$ such that $V = V' \circ \tilde{\pi}_V$.

6.2. **The Bayesian argument revisited.** The information about M described by $\mathcal{M}$ (the σ-field generated by the weak*-topology) is very good information: the weak*-topology separates points on M. That is, *on the basis of his information*, the decision maker is able to distinguish between any two models. In the terminology of sub σ-fields, this is the field associated with the partition generated by the identity mapping. In this case, as we saw above, the "integration over priors" argument works. It is worth recording this formally:

Corollary 2. *If the information about M described by $\mathcal{M}$ and V is linear, then the decision maker's ranking of the acts obeys SEU.*

If the partition is coarser, the argument should be modified along the lines discussed above. In this case, the mapping $\tilde{\pi}_V$ is the familiar conditional expectation operator. Essentially, the idea is that the decision maker would compute a collection of conditional probabilities, one for each element of the partition. Then, he

[6] Whenever V satisfies condition (6.1), there exists, of course, a unique functional V' which makes the diagram commute.

would average these conditionals with the weights that he gives to the corresponding elements of the partition, and SEU would obtain again. Momentarily, we are going to see that this is not guaranteed to work. In fact, we are going to see that the reasons leading to the failure of the argument describe precisely what we are after.

Since we are concerned with establishing conditions for which the integration over prior argument fails, from now on we are going to assume that (unless otherwise stated) the decision maker is described by a nonatomic probability measure P on $(M, \mathcal{M})$.

7. Measurable information

In this section, we are going to study conditions on the information structure which are necessary and sufficient for the "integration over priors" argument to work. We stress that we are concerned with the properties of the partition as a whole, that is of the information structure, rather than with the properties of the individual events making up the partition. In what follows, we only use partitions made up of measurable events, and work modulo sets of measure 0; in particular, when we say partition we mean partition mod 0. So, let $\mathcal{I}$ be a partition of M. The central concept is expressed in the definition below.

Definition 2. ***(Rokhlin*** [27]***)*** *A canonical system of conditional measures associated to the partition $\mathcal{I}$ is a family of measures $\{P_\iota, \iota \in \mathcal{I}\}$, with the following properties*

(i) P_ι is a Lebesgue measure on ι;

(ii) for any $A \in \mathcal{M}$, the set $A \cap \iota$ is measurable in ι for almost all $\iota \in M/\mathcal{I}$ and the function $P_\iota(A \cap \cdot) : M/\mathcal{I} \to \mathbb{R}$ is measurable with

$$P(A) = \int_{M/\mathcal{I}} P_\iota(A \cap \iota) dP'$$

where P' is the pushforward of P under the canonical projection $\pi : M \longrightarrow M/\mathcal{I}$.

In [27], Rokhlin showed that a canonical system of conditional measures exists if and only if there exists a countable family of measurable subsets of $M/\mathcal{I}$ which separate points. In such a case, that is when $M/\mathcal{I}$ is countably separated, the partition $\mathcal{I}$ is called *measurable*.

The idea of measurable partition has an intuitive meaning: loosely, it states that "averages" can be taken, and that things "add up" properly. A bit more precisely (see proof of Theorem 3), the idea is as follows. Given a partition $\mathcal{I}$ of $(M, \mathcal{M}, \mathcal{I})$, consider a system of conditional probabilities $\{P_\iota, \iota \in \mathcal{I}\}$. Then, for each $\varphi \in B(M)$, we have two possible ways of evaluating it: (1) by computing the integral $\int_M \varphi dP$; (2) by computing $\int_{M/\mathcal{I}} \int_\iota \varphi \mid_\iota dP_\iota dP'$, $\iota \in \mathcal{I}$. A partition is called measurable if (a) the expression in (2) makes sense; and (b) $\int_M \varphi dP = \int_{M/\mathcal{I}} \int_\iota \varphi \mid_\iota dP_\iota dP'$, that is the two ways lead to the same evaluation. We can now state our first result about information structures. In the statement, we concisely say "SEU obtains" to mean that the decision maker orders acts in $B(S)$ by means of an expected utility criterion.

Theorem 3. *SEU obtains if and only if the partition $\mathcal{I}$ of M is measurable.*

Notice that if the prior P is such that supp P is contained in a single equivalence class, then (P mod 0) the partition is always measurable. In fact, in such a case, the conditional on supp P coincides with the prior (all others can be specified arbitrarily), and it is straightforward to check that all the conditions in the definition above are satisfied. If we interpret, like we did above, a partition as representing a statistical experiment, this would be a case where the outcome of the experiment is completely uninformative for the decision maker. Motivated by this observation and the result above we give the following definition.

Definition 3. *An information structure is measurable if the associated partition is measurable for every nonatomic prior P on $(M, \mathcal{M})$.*

8. Nonmeasurable information: examples

In this section, we are going to link the notion of nonmeasurable partition with that of Knightian Uncertainty. We will do so by discussing two examples, which have been selected because of their simplicity. Examples more relevant to decision theory will be studied later.

Two popular examples of nonmeasurable partitions are (1) the partition of the torus T^2 by lines of irrational slope α; (2) the partition of the unit interval by means of the equivalence relation $x \sim y$ if and only if $y = x + \alpha \pmod 1$, where, like before, α is a fixed irrational number. We will describe the behavior of the quotient space produced by such equivalence relations first from the viewpoint of the topological properties, and, then, from the viewpoint of the measure space properties. While a discussion of the topological case is not strictly necessary for our purposes, it is convenient to include it, nonetheless. In fact, geometric intuition is simpler in the topological case.

8.0.1. *Topological properties.*

Example (1). To begin, let us consider the unit square along with a partition of it into lines of irrational slope α. Clearly, there are uncountably many of such lines. From the square, obtain the torus T^2 by gluing its sides. The original partition of the square produces a partition of the torus into spirals. Since the original lines had irrational slope, each spiral revolves around the torus without ever meeting itself, and it is easily seen that each spiral is dense for the usual topology of T^2. Now, define an equivalence relation on the torus by declaring two points equivalent if and only if they belong to the same spiral. Clearly, there are uncountably many equivalence classes, that is the quotient has uncountably many points. As a topological space, the quotient is equipped with the finest topology which makes the canonical projection continuous. It is easy to see that such a quotient is not separated (in particular, no point in the quotient can be a closed set as each spiral is dense). In fact, the only closed sets in the quotient are the empty set and the whole quotient. Hence, while as a set it has uncountable many points, as a topological space the quotient behaves as a one-point space (equivalently, the only continuous functions on the quotient are the constants).

Example (2). Consider the mapping from the unit interval into itself given by $f : x \longmapsto x + \alpha \pmod 1$. It is easily checked that for α irrational this mapping has no fixed point. Define an equivalence relation on the unit interval by $x \sim y$ iff $\exists n \in \mathbb{N}$

such that $y = f^n(x)$ (f^n is the n^{th} iterate). One can see that each equivalence class is dense in $[0, 1]$, and that the same conclusion as above about the quotient obtains.

As the reader has certainly noticed, the two examples are essentially the same. Below, we treat them simultaneously.

8.0.2. *Measure space properties.* Given a measure space $(M, \mathcal{M}, P)$ and a partition $\mathcal{I}$ of M, a *basis* for the partition $\mathcal{I}$ is a countable separating family of measurable subsets of $M/\mathcal{I}$. Then, by the theorem of Rokhlin mentioned above, the partition $\mathcal{I}$ is measurable if and only if a basis exists.

Consider again example (1). Now, let T^2 be endowed with the usual measure structure, and let ξ denote the partition of the torus into spirals. It is clear, that each spiral is a measurable subset of T^2. It can be readily checked that the problem of finding a basis for ξ is equivalent to the problem of finding a basis for the partition η of the unit circle T defined as follows. Two points $x, y \in T$ are in the same element of η if and only if $\exists n \in \mathbb{N}$ such that $y = x + n\alpha \pmod 1$. Hence, if we define the map

$$r_\alpha : T \to T \qquad by \qquad x \longmapsto x + \alpha \pmod 1$$

we see that the elements of the partition are precisely the r_α-invariant subsets of T. It is well-known [7], that the map r_α is ergodic, that is every r_α-invariant subset has either measure zero or measure one. This shows that the partition ξ of T^2 is nonmeasurable. Equivalently, the quotient T^2/ξ behaves, when considered as a measure space, as a one-point space, that is the only integrable functions are constant almost everywhere.

8.1. **Ergodicity.** Given a set M and an equivalence relation, $\sim$, on it, subsets of M which are union of equivalence classes are called *saturated* with respect to $\sim$. If M has a measurable structure, $(M, \mathcal{M})$, a measure P on $(M, \mathcal{M})$ is called $\sim$ergodic if the P-measure of any saturated set is either 0 or 1. A nonatomic $\sim$ergodic measure on $(M, \mathcal{M})$ is said to be trivial if its support is contained in a single equivalence class. Otherwise, it is said to be nontrivial, and any saturated set of measure 1 is necessarily the (disjoint) union of measure 0 equivalence classes. Why is this important to us? The reason lays in a theorem of Effros [9], which states that the existence of a nontrivial $\sim$ergodic measure is equivalent to the fact that the quotient space is not countably separated. Then, Rokhlin's theorem implies that there does not exist a canonical system of conditional probabilities (equivalently, the partition is non-measurable).

8.2. **Toward a formalization of Knightian Uncertainty.** In Theorem 3, we saw that if a partition is non-measurable, then the SEU theorem fails, while in this section we have seen some of the characteristic features displayed by non-measurable partitions. How does this relate to Knightian Uncertainty? Informally, the term Knightian Uncertainty refers to situations where there is an inherent inability to assign probabilities to various events. It is intuitively associated to concepts like "insufficient knowledge", "confusion", etc. This is exactly what the idea of nonmeasurable partition conveys. Let us take a closer look. Let us go back to the torus example, and let us consider the topological case first, as intuition is easier in that case. We have seen that the pathological nature of the quotient means that, on the basis of our information (the topology of the torus), we cannot distinguish between any two equivalence classes (spirals). Why is this so? Let us try to distinguish between two spirals. To this end, pick any point on the torus, and consider an open set around that point. Since each spiral is dense in the torus, all the equivalence classes intersect such an open set. Put in another way, each and any spiral has elements having the property described by that open set. Hence, we cannot distinguish among spirals on the basis of such a property. What's more, the denseness of each spiral implies that it is so for any open set: no matter where you place yourself on the torus, on the basis of your information every equivalence class looks like any other. Hence, the inability of distinguishing among them. This is why the quotient looks pathological. It behaves just like the set of reals equipped with the trivial topology: you know that it has lots of points, but you also know that you cannot distinguish among them. The situation is exactly the same if we replace the topological structure with a measure space one. And, if one cannot distinguish between two things, then there is no way of assigning probabilities to them.

9. Uncertainty on the measure zero class

We have seen that non-measurable information structures lead to the inability of assigning probabilities. Now, the issue boils down to whether or not these information structures are mere curiosities. The scope of this and the next section is to show, by means of examples, that they are not.

The first information structure that we consider is not only very natural, but also has an obvious relevance for any theory of decision making. It is the one that partitions the set of models in such a way that two models are equivalent if and only if they are associated to the same collection of measure zero events (in S). Formally, for $\mu, \nu \in \mathcal{P}(S)$, such an information structure is defined by the equivalence relation

$$\mu \mathcal{E} \nu \qquad iff \qquad \mu \ll \nu \qquad and \qquad \nu \ll \mu$$

where $\ll$ stands for absolute continuity, and two measures are equivalent if and only if they are mutually absolutely continuous. Information patterns associated to such an equivalence relation are of the form "The class of measure zero events in Σ is either Φ or Ψ", etc.. Our first result about information structures is stated in the following theorem (see Kechris and Sofronidis [23]).

Theorem 4. *The information structure $\mathcal{E}$ is nonmeasurable.*

The result follows (Appendix B.4) by combining the Spectral Theorem for normal operators with the ergodicity of unitary equivalence . The theorem states that the information structure associated to measure equivalence behaves exactly like the equivalence relation on the torus we studied before. The quotient is not countably separated: on the basis of his information, the decision maker cannot distinguish between any two equivalence classes. If he decides to do something in correspondence to a certain equivalence class, then he must do the same in correspondence to any other equivalence class simply because he cannot distinguish between them. In other words, his behavior must be constant across equivalence classes.

The implication of Theorem 4 is dramatic: Bayesian theory holds only in a very special case. To see this, consider a Bayesian decision maker with a nonatomic probability P on $(M, \mathcal{M})$, and suppose that he tries to integrate over priors in a way that respects his information. Consider the following two disjoint events in M

(a) $Y = \{m \in M \mid m(A) = 0, A \subset S\}$
(b) $Z = \{m \in M \mid m(A) > 0, A \subset S\}$

and let $\psi : M \to R$ be a (hypothetical) bet on M such that ψ takes value y on Y and takes value $z \neq y$ on Z. It is clear that evaluating bets like ψ is crucial to determine the ranking of the bets with domain S. How is our decision maker going to evaluate ψ? Clearly, he has to evaluate the relative likelihood of Y vs Z. Now, we have two possibilities: either supp$P$ intersects both Y and Z (that is, the decision maker is uncertain about whether or not A has nonzero probability) or the decision maker is *a priori* certain that either A has probability zero or that A has nonzero probability (that is, either supp$P \subset Y$ or supp$P \subset Z$). In the first case, the decision maker cannot evaluate ψ: Y and Z are union of equivalence classes from the measure equivalence relation, and the theorem states that such equivalence relation behaves like the one in the torus example. We can say that events like Y and Z are *nonmeasurable* with respect to the decision maker's information. In this case, it is easy to see why the integration over prior argument must fail: since the decision maker cannot assess the likelihood of Y and Z, he cannot take the average of such likelihoods, neither with weights given by P nor in any other way. In the second case, the decision maker can obviously evaluate ψ. If this happens to be the case for each and every event $A \subset S$, then the decision maker can solve his integration over priors problem, and derive a SEU ranking for the acts with domain S. Notice, that this implies that it must be the case that supp$P$ is contained in a single equivalence class. We record this in the following corollary

Corollary 3. *Let d be a Bayesian decision maker whose information structure on the set of models is given by the measure equivalence relation. Then, SEU obtains if and only the decision maker is a priori certain about the class of measure zero events of S.*

In other words, if the only information available to the decision maker regards the class of measure zero events, and if the decision maker is uncertain about this class, then the decision maker cannot be Bayesian.

10. A reconsideration of Ellsberg's Paradox

Here, we consider a continuous version of Ellsberg's three-color urn experiment. The urn is the interval $[0,1]$, which we should think of as partitioned into three subsets, labeled R, B and G. The set of bets is $\{f \mid f : [0,1] \to \mathbb{R}, f \text{ bounded and } \Lambda\text{-measurable}\}$, where Λ is the Lebesgue σ-algebra. The set of models, M, is the set of probability measures on $([0,1],\Lambda)$. The decision maker's information − "the true model belongs to the subset $\Omega \subset \mathcal{P}([0,1])$ such that $\mu(R) = 1/3$ for every $\mu \in \Omega$" − is about the set of models and not about the domain of the bets. We are going to show that if a Bayesian decision maker is given this type of information, then SEU obtains only in one special case. In all others, the information structure associated with the experiment generates an inherent inability to produce a single probability measure on the domain of the bets.

10.1. **An equivalent representation.** Because of the result of the previous section, we focus on Bayesian (nonatomic) decision makers who are *a priori* certain about the class of measure zero events of S. We denote by $[\mu]$ such a class, and by μ one of its representatives. For such decision makers, the theorem below provides a representation equivalent to that of Theorem 2, with the only difference that the subset of probability measures on S is now replaced by the set $G = Aut(\Sigma,\mu)$ of automorphisms of Σ which are nonsingular with respect to μ (see Appendix A.4).

That is, each configuration of the urn is now associated to an element of G (in fact, an equivalence class). The reason for using this version of Theorem 2 is convenience. Later, it will allow us to describe, in a useful way, the information associated with Ellsberg's experiment.[7]

Theorem 5. *Let d be a decision who satisfies the assumptions of Theorem 2, and assume that $M \subset [\mu]$ for some $\mu \in \mathcal{P}(S)$. Then,*

$$f \succcurlyeq_S g \qquad iff \qquad \tilde{V} \circ \tilde{\kappa}(f) \geq \tilde{V} \circ \tilde{\kappa}(g)$$

where
(i) $\tilde{\kappa} : B(S) \to B(G)$ is defined by, for $f \in B(S)$ and $g \in G$,

$$\tilde{\kappa}(f)(g) = \int_S f dg_* \mu$$

with $g_*\mu$ denoting the pushforward of μ under g; and
(ii) $\tilde{V} : B(G) \to \mathbb{R}$.

10.2. **Ellsberg's Paradox.** In the three-color urn experiment, a decision maker faces bets whose domain is an urn containing 90 balls. He is told that of those 30 are red (R) while the remaining are either blue (B) or green (G) in an unknown proportion. As it is well-known, the following violation of the SEU is often observed

$$R \succ B$$

but

$$R \cup G \prec B \cup G$$

In our view, what makes the Paradox actually a paradox is that these decision makers often exhibit indifference between betting on B rather than G. What's more, this is indifference in a very strong sense. One can replace B with G (and vice versa) at any point in decision maker's table of preferences without changing the table itself. It is this symmetric treatment of B and G that prevents us to dismiss those choices as simply incorrect. Such a symmetry is respected in all Ellsberg's decision makers' choices, and, therefore, can hardly be considered an accident.

10.3. **Symmetry in the information.** In Ellsberg's experiment, a decision maker walks into a room, he is offered the bets, and he is told that there are 30 red balls. There is hardly any doubt that the information he receives about the possible configurations of the urn is symmetric in the events B and G. Here, we want to gain a better grasp of this idea of symmetry. In order to explore it, let us begin with a simpler example. We can think of it as some sort of reduced Ellsberg's experiment. Let us suppose that, in addition to the information he is given in the original experiment, the decision maker is told that the number of blue balls is either 20 or 40. Consequently, the number of green balls is either 20 or 40. The symmetry involving B and G is evident.

In our reduced experiment, we have (modulo permutations of the balls) two possible configurations of the urn, each corresponding to an element of the group G.

[7]The description in terms of $Aut(\Sigma, \mu)$ is probably better suited to problems like Ellsberg's experiment as it allows for a wider variety of information patterns. For instance, in Ellsberg's setting, one could order the balls, and give the decision maker information about such an ordering (something like "the 29th ball is red"). This possibility is accounted for when $M = Aut(\Sigma, \mu)$.

Let g_1 and g_2 be such elements. We have already noticed that the two configurations are the same except for the labels, B and G, which are attached to two subsets of the urn. This feature translates into a precise property of our automorphisms g_1 and g_2: there exists a third automorphism $u \in G$ such that $g_1 = u^{-1} g_2 u$. In other words, if two configurations are linked one to another by a relabelling of the domain of the bets, then the corresponding automorphisms are linked one to another by the relation $g_1 = u^{-1} g_2 u$, for some $u \in G$.

In more detail, but with the *caveat* that we are going to define mappings using only the parts the concern us, the situation translates as follows. The domain of the bets is $[0, 1]$, which is equipped with the Lebesgue algebra Λ. The sets R, B and G are placed so that $R = [0, 1/3)$, $B = [1/3, 2/3)$ and $G = [2/3, 1]$. In this section, we have been assuming that the decision maker is *a priori* certain about a measure class on Λ. To fix ideas, suppose that this is the measure class of the Lebesgue measure λ. We take $m_0 = ([0, 1], \Lambda, \lambda)$ as a reference point (i.e., m_0 is associated to the identity in G). In our reduced experiment, we have two models for the decision maker, m_1 and m_2, each of which corresponds to a measure space: $m_1 = ([0, 1], \Lambda, \mu_1)$ is such that $\mu_1(R) = 1/3$, $\mu_1(B) = 2/9$, $\mu_1(G) = 4/9$, while $m_2 = ([0, 1], \Lambda, \mu_2)$ is such that $\mu_2(R) = 1/3$, $\mu_2(B) = 4/9$, $\mu_2(G) = 2/9$. According to Theorem 5, each model corresponds (modulo λ-preserving transformation of $[0, 1]$) to an automorphism of $[0, 1]$. In our case, m_1 and m_2 can be associated to the two automorphisms g_1 and g_2

$$g_1 = \begin{cases} [0,3) & \to & [0,3) \\ [3,5) & \to & [3,6) \\ [5,9] & \to & [6,9] \end{cases} \qquad \text{and} \qquad g_2 = \begin{cases} [0,3) & \to & [0,3) \\ [3,7) & \to & [3,6) \\ [7,9] & \to & [6,9] \end{cases}$$

In fact, the measure associated to model m_i is given by $\lambda_* g_i$, the pushforward of λ under g_i. As observed, m_1 and m_2 are the same except for the labels, B and G, that are attached to two of its subsets. Let u be the automorphisms of the algebra Λ given by

$$u = \begin{cases} [0,3) & \to & [0,3) \\ [3,5) & \to & [7,9] \\ [5,6) & \to & [6,7) \\ [6,9] & \to & [3,6) \end{cases}$$

Then, it is immediate to check that the diagram below commutes

$$\begin{array}{ccc} [0,1] & \xrightarrow{g_1} & [0,1] \\ u \downarrow & & \downarrow u \\ [0,1] & \xrightarrow{g_2} & [0,1] \end{array}$$

That is, $g_1 = u^{-1} g_2 u$. The relation $g_1 = u^{-1} g_2 u$, $u \in G$, defines an equivalence relation on G. It is the orbit equivalence relation (see Appendix A.2) generated by the action by conjugation of G on itself.

10.4. **Treating B and G symmetrically.** Let us begin by recording the content of the discussion above.

Definition 4. *Two models, g_1 and g_2 in G, are the relabelling of one another, $g_1 \sim g_2$, if there exists a $u \in G$ such that $g_1 = u g_2 u^{-1}$.*

Now, let us go back to Ellsberg's experiment. Our decision maker is described by the standard space $(G, \mathcal{B}, P)$, $G = Aut(\Sigma, \mu)$. In addition, he is told that the true model is such that the measure of R is equal to $1/3$. He recognizes that this information is symmetric in B and G, and wants to respect such a symmetry when making his choices. This demands that the set of all possible models be partitioned according to the above equivalence relation. Notice that, given the partition associated to the equivalence relation, the information given to the decision maker takes the form:"the true model belongs to a subset which is the union of a certain number of equivalence classes". Can he be Bayesian?

10.5. **Ergodicity, once again.**

Theorem 6. *The information structure associated to the equivalence relation in Definition 4 is nonmeasurable.*

At this point, we can just repeat what we said in Section 9 following Theorem 4. That is, Theorem 6 leaves us with only two possibilities. We describe them quickly since the argument is exactly the same as in Section 9. The first possibility is that the measure P is trivial, that is the prior is concentrated on a single equivalence class. In terms of Ellsberg's example this means, for instance, that our decision maker believes that the blue balls are either 20 or 40 but he excludes *a priori* that the blue balls might be either 19 or 21 or any other number, even though no one has given him such information. In such a case, the conditional measure exists and coincides with the prior. The decision maker satisfies Savage's axioms on $B(S)$, and behavior of Ellsberg's type cannot be observed. The second possibility is that P is nontrivial, that is its support contains parts of at least two equivalence classes (and, hence, of uncountably many of those). We know that, in such a case, the ergodicity of the information structure means that the quotient $G/\sim$ is not countably separated. In turn, this implies that there does not exist a canonical system of conditional probabilities, and we can find no measure on S able to represent our decision maker. He will not behave in a Savage fashion when it comes to ranking bets in $B(S)$. We can summarize our findings as follows.

Corollary 4. *SEU obtains if and only if the decision maker's prior over G is concentrated on a single equivalence class.*

We will elaborate more on this point in the next section as it casts a new light on the nature of decision making under uncertainty.

11. **Two types of uncertainty**

The equivalence relation we encountered in the previous section is interesting not only because it is associated to Ellsberg's experiment, but especially because it allows us to bring to light an aspect of the problem of decision making that was somehow hidden in Savage's theory. According to the above equivalence relation, two measure spaces (two models for the decision maker) are identified if one can be obtained from the other by means of a (measurable) relabelling of the events. Now, let us spend a moment reflecting on the amount of information embedded in the concept of a measure space. If we are given a measure space, say (S, Σ, μ), we are implicitly told a number of things: for instance, the conditional probability of an event given that another occurred; the relative likelihood of an event with respect to that of any other, etc.. Equivalently, we are told of a class of functions (the measurable functions), of the geometric properties of such a class (the $\mathcal{L}^p$ spaces) and of a mapping (the integral) which evaluates these functions by means of real numbers. In fact, such information permits to reconstruct, in an essentially unique way, the underlying space (S, Σ, μ), and once (S, Σ, μ) is given, such information is immediately available.

If we think of such properties as a collection of abstract relations, we realize that measure spaces that differ only because of the labels assigned to the various events are intrinsically the same, in the sense that all the measure theoretic properties (conditional probabilities, relative likelihoods, etc.) as well as all the geometric properties associated with each measure space (the properties of the function spaces) are the same up to a relabelling of the events. By taking this point of view, we come to realize an important feature of the above equivalence relation, and probably get a better understanding of the problem faced by our decision maker. Within each equivalence class, the geometric properties of the various models were fixed. For our decision maker, to assign a probability to the various models was a matter of assigning probabilities to the names of the events. In contrast, to assign probabilities to models in different classes meant to assign probabilities to different geometric properties. In a sense, within each class it is understood how the world works, and the problem is just which state (which element of the domain of the bets) obtains. The description of how the world works, however, varies as we move from one class to another. Here, the problem is about which description is the correct one. Above, we saw that Savage's theory obtains if and only if the decision maker's prior on the set of models is concentrated on a single equivalence class. We can rephrase this by saying that *Savage's theory implicitly assumes that the decision maker is certain about how the world works.* In all other cases, the Expected Utility Theorem fails.

To us, the distinction between "uncertainty about which state obtains" and "uncertainty about how the world works" is intuitively sound. Also, there seems to be little doubt that both are essential aspects of the problem of decision making under uncertainty. To make things extreme, suppose that in an Ellsberg's type experiment a decision maker is told that the urn contains 30 red balls, 46 blue balls and 14 green balls. Then, such a decision maker is certain about how the world works (the configuration of the urn), yet this does not guarantee that the ball drawn from the urn will be a blue one. Conversely, the decision maker might be told only that the 90-ball urn contains one ball of each color, but might be guaranteed (maybe by a dishonest experimenter) that a blue ball will be drawn. In such a case, the decision maker would be certain about the state, but would know very little about the configuration of the urn. In this respect, we believe that one of the main conceptual contributions of Theorems 1 and 2 has been to uncover the object – the space of models for the decision maker – which makes explicit the presence of these two uncertainties in the process of decision making. Finally, we should like to stress that the intuition that the two uncertainties must be related[8] is formally confirmed in Theorems 1 and 2 by the presence of a mapping linking the domain of the bets to the space of models.

---

[8]The intuition is that a probabilistic assessment as to which state obtains cannot be independent of a probabilistic assessment (if any) of how the world works. For instance, in the example with 46 blue balls, an assessement implying that the probability that a blue ball is drawn from the urn is zero would be patently inconsistent.

12. Comments and complements

12.1. **Knightian Uncertainty.** In this paper, we have formally linked the problem of Knightian Uncertainty to that of the decision maker's information. Building on this, we have then characterized those situations in which the information available relegates the decision maker to a situation of Knightian Uncertainty. For decision makers described by a nonatomic prior on the space of models, our findings can be summarized as follows: *an information structure on M generates Knightian Uncertainty on S if and only if (a) it is nonmeasurable; and (b) the prior on M is not concentrated on a single equivalence class.* The two ingredients identifying Knightian Uncertainty – non-measurable information and non-trivial prior – are both essential. This is intuitive. The second ingredient (non-trivial prior) reveals that Knightian Uncertainty is associated to substantial uncertainty: the decision maker is uncertain about how the world works. The first (non-measurable information) expresses the fact that the decision maker is (in a measurable sense) unable to deal with such uncertainty: a given description of the world looks just like any other, any choice he might make looks just as good as any other. This is what the singularity of the quotient $M/\mathcal{I}$ says.

12.2. **More on singular quotients.** In a set-theoretic sense, singularity of the quotient $M/\mathcal{I}$ means that there is no countable set of properties (Borel sets) which permit to distinguish between two distinct equivalence classes. Anthropomorphizing a bit, suppose that there is an entity who has full knowledge of everything that concerns the set M, and that such an entity agrees to faithfully answer any countable set of questions the decision maker might have about M. With the answers he receives, our decision maker is free to construct any sort of experiment allowing him to test whether or not a certain property is satisfied by some class of models. Singularity of the quotient means that, even in such a case, the decision maker will not be able to distinguish between any two distinct equivalence classes.

12.3. **Two uncertainties and the definition of S.** The distinction between the two uncertainties has nothing to do with the way one defines S. In this paper, we have been interpreting S as the objectively given domain of the acts. We have done so in order to guarantee that the theory be testable. One might then wonder whether or not our distinction would survive if one attributes to S the same universal meaning as in Savage [28]. The point is that, no matter how one defines S, one would still have events in S, and one would still have a decision maker who is uncertain about the relations across them. This is what M (as a space) stands for, and this what is conveyed by Theorems 1 and 2 (where, incidentally, no assumption about the nature of S is made).

In a way, the distinction between the two uncertainties corresponds to that between set and space, that is a set endowed with a certain structure. In a problem of decision making, this structure describes the properties of the environment where decisions take place. Then, Theorems 1 and 2 bring to light not only the intuitive fact that a decision maker might be uncertain about these properties, but also that information about these properties is valuable.

12.4. **Rational Choice.** Savage's axioms are often viewed as providing a satisfactory definition of rational behavior. Since the concept of information has no place in Savage's axioms, it follows that this view implicitly postulates that the rules followed by the decision maker must be independent of the information available. We believe that our results, which state the impossibility of obeying SEU in the presence of certain information structures, cast a serious doubt on this interpretation of Savage's theory. Rather, our results pave the way for a definition of rationality which is a function of the information available to the decision maker. In this regard, it should be noticed that, as a consequence of Theorem 3, any definition that might be proposed must be able to produce models of decision making that are necessarily non-additive in the presence of nonmeasurable information. We will return to this point in the concluding remarks.

13. Subjectively measurable events

When the information about M is described by a non-measurable partition, there exists no system of canonical conditional probabilities (Section 7). In this section, we focus on decision makers who, in the presence of non-measurable information $\mathcal{I}$, use some (necessarily non-canonical) system of conditional probabilities to rank the acts. Any such system,$\{P_\iota\}_{\iota\in\mathcal{I}}$, defines (like in Section 6.1) an operator $\tilde{\pi}_V : \kappa(B(S)) \longrightarrow \mathbb{R}^{M/\mathcal{I}}$, and the necessary condition for the decision maker to respect his information is that the functional V decomposes as seen in the diagram below

$$\begin{array}{ccc} \kappa(B(S)) & \xrightarrow{\tilde{\pi}_V} & \mathbb{R}^{M/\mathcal{I}} \\ & V \quad \searrow & \downarrow V' \\ & & \mathbb{R} \end{array}$$

In such circumstances, however, Theorem 3 tells us that there exist acts $f \in B(S)$ such that either $\tilde{\pi}_V(\kappa(f)) \notin B(M/\mathcal{I})$ (which means that f cannot be measurably evaluated with respect to the decision maker's information) or $\int\limits_M \kappa(f)dP \neq \int\limits_{M/\mathcal{I}} \int\limits_\iota \kappa(f) \mid_\iota dP_\iota dP'$ (which means that f cannot be evaluated consistently). The complement in $B(S)$ of this set, i.e.

$$\Sigma MA = \left\{ f \in B(S) \mid (a)\ \tilde{\pi}_V(\kappa(f)) \in B(M/\mathcal{I});\ (b) \int\limits_M \kappa(f)dP = \int\limits_{M/\mathcal{I}} \int\limits_\iota \kappa(f) \mid_\iota dP_\iota dP' \right\}$$

and its subset

$$\Sigma ME = \{\chi_E \in B(S) \mid \chi_E \in \Sigma MA\}$$

are of special significance in that they describe all those acts in $B(S)$ and events in Σ that can be evaluated measurably with respect to the decision maker's information. In particular, the evaluation of elements in ΣME produces a natural set function on the subset $\{E \in \Sigma \mid \chi_E \in \Sigma MA\}$, which, by a mild abuse of notation, we still denote by ΣME. These observations suggest that we give the following

Definition 5. *A function $f \in B(S)$ is subjectively measurable if $f \in \kappa^{-1} \circ \tilde{\pi}_V^{-1}(B(M/\mathcal{I}))$. An event in Σ is subjectively measurable if its indicator function is subjectively measurable.*

Of course, if $\mathcal{I}$ is measurable, then SEU obtains, every event in Σ belongs to ΣME, every acts in $B(S)$ is in ΣMA and the natural set function on Σ is the "average" measure obtained through the integration over priors theorem (Theorem 3). If $\mathcal{I}$ is non-measurable, ΣME is a proper subset of Σ. Notice that, as it was to be expected, the exact specification of the class of subjectively measurable events depends on the decision maker's preferences since the mapping $\tilde{\pi}_V$ depends on V (Section 6.1). The next proposition, however, tells us that the classes ΣMA and ΣME display some general properties. In addition, the proposition links our notion with that of unambiguous events proposed by [13] and [25]. We denoted by $\mathcal{A}$ the class of unambiguous events in the sense of [13] and [25].

Proposition 1. *The class of subjectively measurable functions is a linear space. Consequently, the class ΣME is a λ-system (in particular, it is nonempty). Furthermore, $\mathcal{A} \subset \Sigma ME$. Finally, for V linear, there exists a natural measure N on ΣME, defined by*

$$N(E) = \int_M \kappa(\chi_E) dP \qquad , \qquad E \in \Sigma ME$$

where P is the measure that defines V.

In general, the inclusion $\mathcal{A} \subset \Sigma ME$ is strict. It is possible, however, to give examples of information structures on M such that the only measurable functions on $M/\mathcal{I}$ are constant (this is the case, for instance, with the torus example). In such a case, it is easily seen that $\mathcal{A}$ corresponds to the intersection over all systems of conditional probabilities of the corresponding classes of subjectively measurable events. In this respect, the notion of [13] and [25] can be viewed as the most restrictive notion of subjectively measurable events compatible with our approach.

Remark 1. *The notion of subjectively measurable events is more general than the one given above in that it does not require the use of systems of conditional probabilities. In fact, any functional V subject to an informational constraint as in (6.1), Section 6.1, produces a mapping $\tilde{\pi}_V : \kappa(B(S)) \longrightarrow \mathbb{R}^{M/\mathcal{I}}$. We can then define the sets ΣMA and ΣME just like above. The inclusion $\mathcal{A} \subset \Sigma ME$ still holds since constant functions on $M/\mathcal{I}$ are always measurable.*

13.1. **Unforeseen contingencies.** Bayesian decision makers cannot measurably evaluate events that are not in ΣME. For this reason, we should expect, for instance, that such events would not be explicitly specified in a contract involving two or more such decision makers. This provides an admittedly tenuous (at this stage, at least) link with the problem of unforeseen contingencies. Focusing on a special case would probably clarify the point. Suppose that the class ΣME does not separate points in S. That is, there exist two points s_1 and s_2 in S such that there exists no $A \in \Sigma ME$ which contains one point but not the other. This means that, on the basis of his information, the decision maker is unable to distinguish between the two states s_1 and s_2. Equivalently, an act $f \in B(S)$ such that $f(s_1) \neq f(s_2)$ cannot be measurably evaluated by the decision maker. In such circumstances, we can define an equivalence relation S by terming two points in S equivalent if they cannot be separated by events in ΣME. The resulting quotient space can then be interpreted as a subjective state space in the spirit of [8]. In a similar vein, acts $f \in B(S)$ which are non-constant on equivalence classes can be reinterpreted as correspondences defined on the quotient in the spirit of [12]. It should be stressed that, by construction, these correspondences do not admit measurable selections.

14. Finite spaces

We have been assuming throughout that the underlying domain of the bets S is uncountable. Strictly speaking, this is essential to our results. The scope of this final section is to point out that the same conclusions can be reached when S is finite. The key is the observation about the nature of singular quotients made in Section 12. In a nutshell, this is how it goes. While we assumed that S was uncountable, at the same time we endowed our decision maker with a tremendous amount of information (the Borel field generated by the weak*-topology). In the finite case, if we endow the decision maker with a more reasonable amount of information (ex. finitely generated algebras) we may reach the same conclusions. The remaining part is an exemplification of this point.

Assume that S has n points. There is a unique measurable structure that makes S into a standard space. This is the one generated by the discrete topology. For such a structure, every act is measurable (and, in fact, continuous). Now, let us put ourselves in the same situation as when we studied Ellsberg's Paradox. The decision maker is a priori certain about the class $[\mu]$ of measure zero events. The measure μ (the representative) is described a vector in the unit simplex in $\mathbb{R}^n$. The set of models is a subset of the simplex. It can be equivalently described as follows. To begin, observe that, for any $1 \leq p \leq \infty$, all the $\mathcal{L}^p(S, \Sigma, \mu)$ spaces are equivalent, and can be thought of as the Hilbert space $\mathbb{R}^n$ with the usual scalar product. The measure μ defines a linear functional (an integral) on $\mathbb{R}^n$ by $\langle f, \mu \rangle$, $f \in \mathbb{R}^n$. Such a functional gives an ordering for the elements of $\mathbb{R}^n$, which are the acts offered to the decision maker. If ν is another measure in $[\mu]$, then ν gives rise to another ordering of the bets, the one given by $\langle f, \nu \rangle$, $f \in \mathbb{R}^n$. Fixing μ, one can represent each measure in the set of models by means of a matrix on $\mathbb{R}^n$. If ν is such a measure, then the corresponding matrix A is defined (modulo equivalence) by $\nu = A\mu$. Hence, the set of models is identified to a subset of the matrices on $\mathbb{R}^n$.

The set of models is partitioned in a way that two models (matrices) are in the same element of the partition if they display the symmetry property discussed above. Here, that means that two models, A and B, are equivalent, $A \sim B$, if and only if there exists a (positive) unitary matrix, U, such that $A = U^*BU$ (a matrix U is a unitary matrix if $U^*U = I$). That is, if and only if A and B are unitarily equivalent. Now, our question is: Does there exist a countable set of properties allowing the decision maker to distinguish between any two equivalence classes?

The answer is provided by an elementary result in Linear Algebra. In fact, two matrices are unitarily equivalent if and only if they have the same set of eigenvalues. It follows at once (think of the Euclidean topology of $\mathbb{R}^n$) that there exists a countable set of properties that distinguishes between any two equivalence classes. In particular, this is certainly the case if the set of models is equipped with a standard Lebesgue structure.

So, it seems that no Knightian Uncertainty may arise in the finite case. In actual situations, however, there is no general argument guaranteeing that a decision maker might be able to distinguish among countably many events. If we admit this, things might turn out to be different. It might happen that, say, at least n properties are needed to distinguish among equivalence classes but the decision maker can distinguish, on the basis of his information, only among k of those for $k < n$. In such a case, we would be back to singular quotients. As a simple example, consider an Ellsberg's type experiment with M being the two-point set $\{\mu_1 = (1/3, 2/9, 4/9), \mu_2 = (1/3, 4/9, 2/9)\}$ and a decision maker who has the trivial algebra $\{\varnothing, M\}$ on M.

15. Conclusion

We have seen that a general process of decision making involves uncertainty about both the set of states and the set of models. As a consequence, a decision maker can obtain valuable information about both sets. We showed that information structures on the set of models split into two categories: those for which the SEU theorem obtains and those that generate, for Bayesian decision makers, an inability to assign probabilities to many events. Situations of Knightian Uncertainty were then identified with the presence of such information structures. The next step is to introduce functionals on $B(\mathcal{M})$ – the bounded measurable functions on $(M, \mathcal{M})$ – which (a) are linear whenever the information structure on M is measurable; and (b) maintain, in some sense, the spirit of linearity whenever the information on M is nonmeasurable. Once this is accomplished, models of decision making (necessarily non-linear) that emerge in case (b) would appear – by analogy with the interpretation associated with Savage's theory in risky situations – as a "rational choice" response to situations of Knightian Uncertainty.

References

[1] Aliprantis C.D. and K.C. Border (1994), Infinite Dimensional Analysis, Springer-Verlag.
[2] Amarante M. (2005), Analogical reasoning in decision processes, Columbia University, mimeo.
[3] Arrow K.J. (1974), Exposition of a theory of choice under uncertainty, in Essays in the Theory of Risk-Bearing, North Holland.
[4] Bewley T. (2002), Knightian decision theory: Part I, *Decision in Economics and Finance* **25**, 79-110.
[5] Billingsley P. (1995), Probability and Measure, Wiley.
[6] Conway, J. B. (1990), A Course in Functional Analysis, Springer-Verlag GTM.
[7] Cornfeld I.P., S.V. Fomin and Y.G. Sinai (1982), Ergodic Theory, Springer-Verlag.
[8] Dekel E., B. Lipman and A. Rustichini (2001), Representing Preferences with a Unique Subjective State Space, *Econometrica* **69**, 891-934.
[9] Effros E.G. (1965), Transformation Groups and C*-algebras, *The Annals of Mathematics* **81**, 38-55.
[10] Ellsberg, D. (1961), Risk, Ambiguity and the Savage Axioms, *Quarterly Journal of Economics* **75**, 643-69.
[11] Fishburn P.C. (1973), A Mixture-Set Axiomatization of Conditional Subjective Expected Utility, *Econometrica* **41**, 1-25.
[12] Ghirardato P. (2001), Copying with Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty, *Economic Theory* **17**, 247-76.
[13] Ghirardato P., F. Maccheroni and M. Marinacci (2004), Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory* **118**, 133-173.
[14] Gilboa I. and D. Schmeidler (1989), Maxmin expected utility with a non-unique prior, *Journal of Mathematical Economics* **18**, 141-53.
[15] Gilboa I. and D. Schmeidler (2001), A Theory of Case-Based Decisions, Cambridge University Press.
[16] Gilboa I. and D. Schmeidler (2003), Inductive Inference: An Axiomatic Approach, *Econometrica* **71**, 1-26.
[17] Harrington, L. A., A. S. Kechris and A. A Louveau (1990), Glimm-Effros dichotomy for Borel equivalence relations, *Journal of the American Mathematical Society* **3**, 903–928.

[18] Hjorth G. (2000), Classification and orbit equivalence relations, American Mathematical Society.
[19] Kahneman D. and A. Tversky (1979), Prospect theory: An analysis of decisions under risk, *Econometrica* **47**, 313-27.
[20] Karni E. (2004), Subjective Expected Utility Theory without States of the World, The Johns Hopkins University, mimeo.
[21] Kechris, A. (1995), Classical Descriptive Set Theory, Springer-Verlag GTM.
[22] Kechris, A. (1999), New Directions in Descriptive Set Theory, *The Bulletin of Symbolic Logic* **5**, 161-74.
[23] Kechris, A. and N.E. Sofronidis (2001), A strong generic ergodicity property of unitary and self-adjoint operators, *Ergodic Theory and Dynamical Systems* **21**, 1459-79.
[24] Marinacci M. (1997), A simple proof of a basic result for multiple priors, University of Toronto, mimeo.
[25] Nehring K. (2001), Ambiguity in the context of probabilistic beliefs, UC Davis, mimeo.
[26] Phelps R.R. (1966), Lectures on Choquet Theorem, Van Nostrand.
[27] Rokhlin, V.A. (1949), On the Fundamental Ideas of Measure Theory, Translations of the AMS, 71.
[28] Savage L.J. (1972), The Foundations of Statistics, Dover.
[29] Schmeidler D. (1989), Subjective probability and expected utility without additivity, *Econometrica* **57**, 571-87.
[30] Srivastava, S.M. (1998), A Course on Borel Sets, Springer-Verlag GTM.

APPENDICES

A. Background material

A.1 Standard Spaces

A *Polish space*, (X, τ), is a separable, completely metrizable topological space. Given the topology τ on X, the Borel σ–field is the one generated by the closed sets. A *Standard Borel space* is a Polish space stripped down to its Borel structure.

Let X and Y be two measurable spaces. A mapping $X \to Y$ is called a Borel isomorphism if it is a bijection and is bimeasurable. An important and well-known fact about standard Borel spaces is stated in the following theorem (see [30], Theorem 3.3.13)

Borel isomorphism theorem Any two uncountable standard Borel spaces are Borel isomorphic.

A Standard Borel space along with a nonatomic measure is a called a *Standard Lebesgue space*. We recall that a measurable set in a Standard Lebesgue space is a set which differs from a Borel set by a set of measure zero.

A.2 Group actions

A group G is a set along with a law of composition which is associative, has a unit element e, and each element in G has an inverse. Let G be a group and X be a set. An *action* of G on X (on the left) is a mapping $G \times X \to X$, written as $(g, x) \longmapsto gx$ $(g \in G, x \in X)$, such that for all $g, h \in G$ and $x \in X$

$$(gh)x = g(hx) \quad and \quad ex = x$$

Let $x \in X$. The set $Gx = \{y \in G \mid \exists g \in G \quad st \quad y = gx\}$ is called the orbit of x under G. Since G is a group, the relation on G

$$x \sim y \qquad iff \qquad \exists g \in G \quad st \quad y = gx$$

is an equivalence relation on G. It is called the *orbit equivalence relation* generated by the action of G on X.

The two examples of group actions that we are going to be using are the following. Let G be a group, and H be a subgroup of G. Then,

(i) The action by conjugation of H on G is the mapping $H \times G \to G$ defined by

$$(h, g) \longmapsto hgh^{-1}$$

(ii) The action by (left) translation of H on G is the mapping $H \times G \to G$ defined by

$$(h, g) \longmapsto hg$$

A.3 Polish groups

If G is a topological space, then G is a topological group if the group operation is continuous [the law of composition is a mapping $G \times G \to G$, $G \times G$ endowed with the product topology].

If G and X are topological (measurable) spaces, the action is continuous (measurable) if the mapping $G \times X \to X$ above is continuous (measurable).

A Polish group is a topological group such that its topology is Polish and the group operation is continuous in such a topology.

Let G be a Polish group and X be a Polish space. Becker and Kechris have shown that if G acts in a Borel way on X, then there are equivalent Polish topologies on G and X for which the action is continuous.

A.4 The Polish group $G = Aut(\Sigma, \mu)$

Consider the measure space (S, Σ, μ). An automorphism of the measure algebra Σ is a bijection on Σ. If g is one such an automorphism, g is said to be *nonsingular* if the pushforward of μ under g is absolutely continuous with respect to μ. In symbols, $g_*\mu \ll \mu$. We denote by $G = Aut(\Sigma, \mu)$ the set of automorphisms which are nonsingular with respect to μ. It is immediate to verify that G is a group with the operation of composition of mappings.

Let (S, Σ, μ) be a measure space, and $G = Aut(\Sigma, \mu)$. Denote by $\mathcal{L}^2(\mu)$ the Hilbert space of square integrable functions on (S, Σ, μ).

There is a natural isomorphism (see Kechris [21], Ch. 17) of the set of nonsingular automorphisms G into the set of positive linear isometries of $\mathcal{L}^2(\mu)$, $g \longmapsto T_g$, where T_g is defined by

$$T_g(f) = f \circ g^{-1} \cdot \left(\frac{dg_*^{-1}\mu}{d\mu} \right)^{1/2} \qquad , \qquad f \in \mathcal{L}^2(\mu)$$

with the expression in brackets denoting the Radon-Nikodym derivative. A topology is defined on G by using the isomorphism and observing that the strong and the weak operator topology coincide when restricted to the positive linear isometries of $\mathcal{L}^2(\lambda)$. Such a topology has some remarkable properties. Here are a few that we will be using in the remaining proofs.

(i) G is a Polish and, hence, a Borel space;

(ii) G is a Polish group;

(iii) An automorphism $g \in G$ is μ-preserving is for every integrable function f on S, $\int f d\mu = \int f dg_*\mu$. Let $PR(\mu)$ denote the group of μ-preserving automorphisms of G. Then, $PR(\mu)$ is a Polish subgroup.

B. Proofs of results in the main text

B.1 Theorem 1

The following observation is useful to understand the nature of the result contained in Theorem 1. Recall that the dual B^* of $B(S)$ is (identified with) the set of all finitely additive measures on Σ. For $M \subset B^*$, the mapping κ appearing in the statement of Theorem 1 is simply the canonical mapping $B(S) \longrightarrow A(M)$, the set of continuous affine mappings on M.

Proof of Theorem 1. 0. Define κ by $f \longmapsto \kappa(f)$ where

$$\kappa(f)(\mu) = \int f d\mu \qquad , \qquad \mu \in \mathcal{C} \subset ba_1(\Sigma)$$

Next, set $M = \mathcal{C}$ and let $\mathcal{M}$ be the Borel σ-algebra generated by the weak*-topology $\sigma(ba_1(\Sigma), B(S,\Sigma))$. Then, the first part follows from GMM's result by observing that (a) if $f, g \in B(S)$ are such that $\kappa(f) = \kappa(g)$, then $\alpha(f) = \alpha(g)$; and (b) $\kappa(f) : M \to \mathbb{R}$ is continuous for the weak*-topology $\sigma(ba_1(\Sigma), B(S,\Sigma))$ (hence, it is measurable for the Borel σ-algebra generated by that topology; see [13] and [2]).

1. κ is sup-norm to sup-norm continuous:

This is a consequence of the following inequality. For $f, g \in B(S)$

$$\begin{aligned} \|\kappa(f) - \kappa(g)\|_\infty &= \sup_{m \in M} \left| \int f dm - \int g dm \right| \leq \sup_{m \in M} \int |f - g| \, dm \\ &\leq \sup_{m \in M} \int \sup_{s \in S} |f - g| \, dm = \|f - g\|_\infty \end{aligned}$$

2. κ is linear: immediate.

Next, observe that any element in $\kappa(B(S))$ is, by construction, a weak*-continuous linear functional on B^* which is restricted to a subset M of B^*.

3. κ is an open mapping:

We need to show that $\kappa(B(S))$ is a norm-closed linear subspace (hence, Banach) of the Banach space $B(M)$: Let $\{\psi_n\}_{n \in \mathbb{N}} \subset \kappa(B(S))$ be such that $\psi_n \longrightarrow \psi$ in the sup norm. Then ψ is the uniform limit of weak*-continuous functions, hence it is weak*-continuous. Since all ψ_n are affine so is ψ. Hence, ψ is a weak*-continuous affine functional and (by definition) there exists an $f \in B(S)$ such that $\psi(m) = \displaystyle\int f dm$. That is, $\psi \in \kappa(B(S))$. Finally, since κ is a continuous linear surjective mapping between Banach spaces, then it is an open mapping by the Open Mapping Theorem.

Next, we want to show that V is a continuous functional for the sup-norm topology on $B(M)$. We divide the proof in several steps.

4. The functions max : $\kappa(B(S)) \longrightarrow \mathbb{R}$ and min : $\kappa(B(S)) \longrightarrow \mathbb{R}$ are both continuous for the sup norm topology on $\kappa(B(S))$:

Consider the mapping $H : \kappa(B(S)) \times M \longrightarrow \mathbb{R}$ (product topology on $\kappa(B(S)) \times M$) defined by $H(\psi, m) = \psi(m)$. Then, for any net $\{(\psi_\alpha, m_\alpha)\} \subset \kappa(B(S)) \times M$

such that $(\psi_\alpha, m_\alpha) \longrightarrow (\psi, m)$ we have

$$\begin{aligned} |H(\psi_\alpha, m_\alpha) - H(\psi, m)| &= |\psi_\alpha(m_\alpha) - \psi(m)| \\ &\leq |\psi_\alpha(m_\alpha) - \psi(m_\alpha)| + |\psi(m_\alpha) - \psi(m)| \\ &\leq \sup_{m_\alpha \in M} |\psi_\alpha(m_\alpha) - \psi(m_\alpha)| + |\psi(m_\alpha) - \psi(m)| \\ &= \|\psi_\alpha - \psi\|_\infty + |\psi(m_\alpha) - \psi(m)| \end{aligned}$$

Hence, $\psi_\alpha \longrightarrow \psi$ (sup norm) in $\kappa(B(S))$ and ψ continuous on $(M, \sigma(M, B(S, \Sigma))$ imply that H is continuous. Now, the compactness of M and the continuity of H imply, by an application of the Maximum Theorem (see [1]), that both max and min are continuous.

5. The functions $\max \circ \kappa : B(S) \longrightarrow \mathbb{R}$ and $\min \circ \kappa : B(S) \longrightarrow \mathbb{R}$ are both continuous for the sup norm topology on $B(S)$: immediate.

Let $\alpha : B(S) \longrightarrow [0, 1]$ be the mapping appearing in GMM's theorem. Then, whenever it is defined (i.e., whenever $\kappa(f)$ is non-constant), $\alpha(f)$ is given by the expression

$$\alpha(f) = \frac{I(f) - \max_{m \in M} \kappa(f)(m)}{\min_{m \in M} \kappa(f)(m) - \max_{m \in M} \kappa(f)(m)}$$

6. α is sup-norm continuous on an open dense subset of $B(S)$:

Whenever it is defined α is sup-norm continuous. α it is not defined on the set

$$N = \{f \in B(S) : \kappa(f) = \text{constant}\}$$

Let $C1$ denote the one-dimensional linear subspace of $\kappa(B(S))$ consisting of the constant functions. Then, $N = \kappa^{-1}(C1)$ and (i) N is closed because $C1$ is closed and κ is continuous; (ii) N has empty interior because κ is an open mapping and $C1$ has empty interior.

7. It follows that α has a continuous extension to the whole $B(S)$.

8. Define $\tilde{\alpha} : \kappa(B(S)) \longrightarrow \mathbb{R}$ by means of the diagram

$$\begin{array}{ccc} B(S) & \xrightarrow{\kappa} & \kappa(B(S)) \\ & \alpha \searrow & \downarrow \tilde{\alpha} \\ & & [0,1] \end{array}$$

This can be done because of 0. above. Now, $\tilde{\alpha}$ is sup-norm continuous because α is and κ is an open mapping.

9. It follows that $V : \kappa(B(S)) \longrightarrow \mathbb{R}$ in the first part of the theorem is defined by

$$V(\psi) = \tilde{\alpha}(\psi) \min \psi + (1 - \tilde{\alpha}(\psi)) \max \psi$$

and is sup-norm continuous by 8. and 4. above. □

B.2 Theorem 2

We recall that, given two measurable spaces, $(Y, \mathcal{Y})$ and $(Y', \mathcal{Y}')$, a mapping $\kappa : B(Y) \to B(Y')$ is called normal if

$$f_n \nearrow f \Longrightarrow \kappa(f_n) \nearrow \kappa(f), \qquad n \in \mathbb{N}$$

Proof of Theorem 2. From Theorem 1, we know that M is a weak*-compact subset of $(ba_1, \sigma(ba_1, B(S)))$. From [13] (Sec. B.3), Axiom 7 is equivalent to the property that all the measures in Theorem 1 are countably additive. That is, $M \subset \mathcal{P}(S)$, endowed with the topology that inherits as a subset of $(ba_1, \sigma(ba_1, B(S)))$. If we replace this topology with $(\mathcal{P}(S), \sigma(\mathcal{P}(S), C_b(S)))$, then M remains compact because the new topology is weaker than the original one. Now, S Polish implies that $\mathcal{P}(S)$ is Polish. Hence, M, being closed, is Polish as well.

Next, define κ as in the statement of the Theorem. Clearly, κ is linear. Since all the probabilities are countably additive, the Monotone Convergence Theorem implies that κ is normal. It remains to check that *range* $\kappa \subset B(M)$. Let $E \in \Sigma$, and let χ_E denote its indicator function. Then, $\kappa(\chi_E)$ is obviously bounded and it is well-known that $\kappa(\chi_E)$ is measurable for the Borel σ-algebra generated by $\sigma(\mathcal{P}(S), C_b(S))$ ([1], Lemma 12.14). If $h \in B(S)$ is a simple function, then it can be written as a linear combination of indicator functions, and linearity of κ implies that $\kappa(h)$ is measurable. If $f \in B(S)$ is any function, then there exists a sequence of simple functions $\{f_n\} \subset B(S)$ such that $f_n \nearrow f$, and normality of κimplies that $\kappa(f)$ is measurable. Finally, define V as in Theorem 1. □

B.3 Theorem 3

Proof of Theorem 3. If $\mathcal{I}$ is measurable, then by Rokhlin [27] there exists a canonical system of conditional probabilities $\{P_\iota\}_{\iota\in\mathcal{I}}$. By using Rokhlin's definition, it is straightforward to check that for every $\varphi \in B(M)$, we have

$$\int\limits_M \varphi dP = \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \varphi \mid_\iota dP_\iota dP'; \quad \iota \in \mathcal{I}, \quad \varphi \in B(M)$$

In particular, the function $(\int\limits_{\iota} \varphi \mid_\iota dP_\iota)_{\iota\in\mathcal{I}}$ is measurable. This means that the conditional expectation operator $T_{\mathcal{I}} : \varphi \longrightarrow (\int\limits_{\iota} \varphi \mid_\iota dP_\iota)_{\iota\in\mathcal{I}}$ satisfies the condition *range* $T_{\mathcal{I}} \subset B(M/\mathcal{I})$, and

$$\int\limits_M \varphi dP = V(\varphi) = \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \varphi \mid_\iota dP_\iota dP' = V' \circ T_{\mathcal{I}}(\varphi) \qquad , \qquad \forall \varphi \in B(M)$$

That is, both conditions (*) and (**) in Section 6 are satisfied. Finally, recall that the decision maker orders acts in $B(S)$ by means of $I = V' \circ T_{\mathcal{I}} \circ \kappa = V \circ \kappa$, and that κ is linear. Now, the statement follows by applying Riesz representation theorem.

Conversely, let $\mathcal{I}$ be nonmeasurable, and let $\{P_\iota\}_{\iota\in\mathcal{I}}$ be a system of conditional probabilities, with each P_ι a Lebesgue measure on ι. By Rokhlin's theorem, $\{P_\iota\}_{\iota\in\mathcal{I}}$ cannot be canonical. Hence, $\exists \varphi \in B(M)$ such that either $T_{\mathcal{I}}(\varphi) : M/\mathcal{I} \longrightarrow \mathbb{R}$ is nonmeasurable or $\int\limits_M \varphi dP \neq \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \varphi \mid_\iota dP_\iota dP'$. Either way, at least one of the conditions (*) and (**) of Section 6 is violated. If such a φ belongs to *range* $\kappa(B(S))$, we are done. Now, we are going to show that *range* $\kappa(B(S))$ necessarily contains at least one such a φ.

To begin, observe that the (non-canonical) system of conditional probabilities $\{P_\iota\}_{\iota\in\mathcal{I}}$ defines an operator $\tilde{T} : B(M) \longrightarrow \mathbb{R}^{M/\mathcal{I}}$ by

$$\psi \mapsto \tilde{T}(\psi) \qquad where \qquad \tilde{T}(\psi)(\iota) = \int\limits_M \psi dP_\iota$$

Also, observe that $\text{supp} P_\iota \subset \iota$.

Let

$$\Theta = \left\{ \psi \in B(M) \mid (a)\ \tilde{T}(\psi) \in B(M/\mathcal{I});\ (b) \int\limits_M \psi dP = \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \psi \mid_{\iota} dP_{\iota} dP' \right\}$$

It is easily checked that Θ is a linear subspace.

Now, let $\{\psi_n\}_{n\in\mathbb{N}}$ be a sequence in Θ;

CLAIM: If either $\psi_n \nearrow \psi \in B(M)$ or $\psi_n \searrow \psi \in B(M)$, then $\psi \in \Theta$.

Proof of the claim: Let $\psi_n \nearrow \psi \in B(M)$.

(a) By the Dominated Convergence Theorem (DCT), $\forall P_\iota$ we have $\int\limits_M \psi_n dP_\iota \nearrow \int\limits_M \psi dP_\iota$, that is $\tilde{T}(\psi_n) \nearrow \tilde{T}(\psi)$. Hence, $\tilde{T}(\psi)$ is a pointwise limit of measurable functions, and hence measurable. Moreover, since $\psi \in B(M)$, $\tilde{T}(\psi)$ is bounded, i.e. $\tilde{T}(\psi) \in B(M/\mathcal{I})$.

(b) Observe that

$$\begin{aligned}
\int\limits_M \psi dP &= \lim_{n\to\infty} \int\limits_M \psi_n dP && \textit{(by the DCT and } \psi \in B(M)\textit{)} \\
&= \lim_{n\to\infty} \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \psi_n dP_\iota dP' && \text{(because } \psi_n \in \Theta\text{)} \\
&= \lim_{n\to\infty} \int\limits_{M/\mathcal{I}} \tilde{T}(\psi_n) dP' \\
&= \int\limits_{M/\mathcal{I}} \tilde{T}(\psi) dP' && \textit{(by (a) and the DCT)}
\end{aligned}$$

$$= \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \psi \mid_{\iota} dP_\iota dP'$$

which completes the proof for the case $\psi_n \nearrow \psi$. The other case is similar.

Now suppose, by the way of contradiction, that *range* $\kappa(B(S)) \subset \Theta$. Let K denote the set of (continuous) convex functions on M. Then, if $\gamma \in K$ there exists ([26], p. 19) $\{\zeta_n\}_{n\in\mathbb{N}} \subset A(M) \subset$ *range* $\kappa(B(S))$ such that $\zeta_n \nearrow \gamma$. By the above claim, $\gamma \in \Theta$, that is $K \subset \Theta$. Since Θ is a linear space, it follows that $K - K \subset \Theta$. By the Stone-Weierstrass theorem, $K - K$ is uniformly dense in $C(M)$, the set of continuous functions on M.

Since M is a metric space, for any closed set $A \subset M$, there exists ([1], Corollary 3.14) $\{\lambda_n\}_{n\in\mathbb{N}} \subset C(M)$ such that $\lambda_n \searrow \chi_A$, where χ_A denotes the indicator function of A. Since $K - K$ is uniformly dense in $C(M)$, for each $n \in \mathbb{N}$, there exists $\{h_{n_k}\}_{k\in\mathbb{N}} \subset K - K$ such that $h_{n_k} \to \lambda_n$ uniformly as $k \to \infty$.

Now, let $k_0 \in \mathbb{N}$ be such that $\forall k \geq k_0$

$$\lambda_0(m) - 1 < h_{0_{k_0}}(m) < \lambda_0(m) + 1 \qquad \forall m \in M$$

Then, the function

$$g_0 = h_{0_{k0}} + 2$$

is in Θ because Θ is a linear space, and satisfies

$$\lambda_0(m) + 1 < g_0(m) < \lambda_0(m) + 3 \qquad \forall m \in M$$

Next, let $k_1 \in \mathbb{N}$ be such that $\forall k \geq k_1$

$$\lambda_1(m) - \frac{1}{3} < h_{1_{k1}}(m) < \lambda_1(m) + \frac{1}{3} \qquad \forall m \in M$$

Then, $g_1 = h_{1_{k1}} + \frac{2}{3} \in \Theta$ and satisfies

$$\lambda_1(m) + \frac{1}{3} < g_1(m) < \lambda_1(m) + 1 \qquad \forall m \in M$$

Moreover, for every $m \in M$, we have

$$g_1(m) < \lambda_1(m) + 1 \leq \lambda_0(m) + 1 < g_0(m)$$

Inductively, define

$$g_n = h_{n_{kn}} + \frac{2}{3^n}$$

Then, $\{g_n\}_{n\in\mathbb{N}} \subset \Theta$, $g_{n+1}(m) < g_n(m)$ $\forall m \in M$, and

$$\sup_{m\in M} |g_n(m) - \lambda_n(m)| < \frac{1}{3^{n-1}}$$

Now, the inequality

$$|g_n(m) - \chi_A(m)| \leq |g_n(m) - \lambda_n(m)| + |\lambda_n(m) - \chi_A(m)|$$

shows that $g_n \searrow \chi_A$. [Notice that $g_n(m) > \lambda_n(m) + \frac{1}{3^n} \geq \chi_A(m)$]

By the above claim, we then have $\chi_A \in \Theta$ for any closed set $A \subset M$.

Next, observe that:

(i) $\chi_M \in \Theta$ because the function $1 \in A(M) \subset \Theta$;

(ii) if $\chi_A, \chi_B \in \Theta$ and $A \subset B$, then $\chi_{B\setminus A} = \chi_B - \chi_A \in \Theta$ because Θ is a linear space;

(iii) if $A_n \nearrow A$ and $\{\chi_{A_n}\} \subset \Theta$, then $\chi_{A_n} \nearrow \chi_A$ and $\chi_A \in \Theta$ by the claim above.

Hence, we conclude that $\mathcal{A} = \{A \in \mathcal{M} \mid \chi_A \in \Theta\}$ is a Dynkin system, which contains all closed sets. Hence, $\mathcal{A} = \mathcal{M}$ (the Borel σ-algebra generated by the topology on M).

But now, it follows that Θ contains all the simple functions (because Θ is a linear space) and since $\{\psi_n\} \subset \Theta$ and $\psi_n \nearrow \psi$ imply $\psi \in \Theta$, we can conclude that $\Theta = B(M)$, a contradiction. □

B.4 Theorem 4

Let $\mathcal{E}$ and $\mathcal{F}$ be two equivalence relations on Polish spaces, X and Y, respectively. Following Kechris [22], we say that $\mathcal{E}$ Borel reduces to $\mathcal{F}$ if there exists a Borel map $f : X \to Y$ such that

$$x\mathcal{E}y \qquad \Longleftrightarrow \qquad f(x)\mathcal{F}f(y)$$

At once, this implies that there exists an embedding $X/\mathcal{E} \to Y/\mathcal{F}$. We say that $\mathcal{E}$ and $\mathcal{F}$ are Borel equivalent, $\mathcal{E} \sim_B \mathcal{F}$, if $\mathcal{E}$ Borel reduces to $\mathcal{F}$ and $\mathcal{F}$ Borel reduces to $\mathcal{E}$.

Proof of Theorem 4. Let $\mathcal{U}$ denote unitary equivalence of normal operators on a separable Hilbert space. By the Spectral theorem (see, for instance, [6] pp. 293-97), every normal operator is associated to a measure class on a Polish space, and two operators are unitarily equivalent if and only if they are associated to the same measure class. If $\mathcal{E}$ denotes measure equivalence (mutual absolute continuity) of (Borel) probability measures on an uncountable Polish space, then we have $\mathcal{U} \sim_B \mathcal{E}$. The proof is completed by collecting the following facts: (a) The action by conjugation of $\mathcal{U}$ on itself is properly generically ergodic (Hjorth [18], Definition 3.1, Theorem 7.7 and Corollary 7.8); this, in turn, implies that (b) there exists a non-trivial non-atomic measure for the equivalence relation produced by such an action (Harrington-Kechris-Loveau [17], Theorem 1.1). Now, Effros' theorem implies that the associated partition is non-measurable. □

Remark 2. *Kechris and Sofronidis* [23] *have shown that a stronger property holds. Namely, the action by conjugation of $\mathcal{U}$ on itself is turbulent (see* [23]*), a fact not needed in this paper.*

B.5 Proof of Theorem 5

The proof of the theorem is based on a result of Effros-Mackey (Lemma 3 below), and on two lemmata, which we prove next.

Define an equivalence relation, $\approx$, on G by (the groups G and $PR(\mu)$ are defined in **A.4**)

$$g_1 \approx g_2 \qquad iff \qquad g_{1*}\mu = g_{2*}\mu$$

Lemma 1. *$g_1 \approx g_2$ if and only if there exists a $b \in PR(\mu)$ such that $g_1 = g_2 \circ b$.*

Proof. Let b be a μ-preserving automorphism of Σ, and let b be such that $g_1 = g_2 \circ b$. Then, $\forall A \in \Sigma$, we have

$$g_{1*}\mu(A) = \mu(g_1^{-1}(A)) = \mu(b^{-1} \circ g_2^{-1}(A)) = \mu(g_2^{-1}(A)) = g_{2*}\mu(A)$$

Conversely, assume $g_{1*}\mu = g_{2*}\mu$. Define $b = g_2^{-1} \circ g_1$. Since both g_1 and g_2 are automorphisms of Σ, so is b. We have only to show that $b \in PR(\mu)$. $\forall E \in \Sigma$, we have

$$b_*\mu(E) = \mu(g_2^{-1} \circ g_1(E)) = g_{2*}\mu(g_1(E)) = g_{1*}\mu(g_1(E)) = \mu(E)$$

□

The lemma says that the equivalence relation $\approx$ on G is nothing other than the orbit equivalence relation produced by the action by right translation of $PR(\mu)$ on G (a fact, that we will be using shortly). Because of this, we denote the quotient $G/\approx$ by $G/PR(\mu)$. Note, also, that the lemma establishes a bijection between $G/PR(\mu)$ and $[\mu]$.

Lemma 2. *$PR(\mu)$ is a closed subgroup of G.*

Proof. Let $E \in \Sigma$, and let χ_E denote the indicator function of E. If $g, \gamma \in PR(\mu)$, then

$$\begin{aligned} \|T_\gamma \chi_E - T_g \chi_E\| &= \int |\gamma \circ \chi_E - g \circ \chi_E| \, d\mu \\ &= \mu(\gamma(E) \triangle g(E)) \end{aligned}$$

where $\triangle$ denotes the symmetric difference. Hence, if $\{\gamma_n\}$ is a sequence in $PR(\mu)$ converging to $g \in G$, by definition the sequence $\left\|T_{\gamma_n}\chi_E - T_{\gamma_m}\chi_E\right\| \to 0$ for each $E \in \Sigma$, and, by the above, so does the sequence $\mu(\gamma_n(E)\triangle\gamma_m(E))$. It follows that g preserves μ. □

Lemma 3. *$G/PR(\mu)$ is standard.*

Proof. By a theorem of Effros-Mackey (see Srivastava [30], p. 196), we have that the partition generated by the action by right translation of $PR(\mu)$ on G admits a Borel transversal. In turn, this implies (Srivastava [30], p. 197) that the quotient $G/PR(\mu)$ is standard. □

Proof of Theorem 5. From Theorem 2 and the assumptions contained in the statement, we know that the diagram below commutes

$$\begin{array}{ccc} B(S) & \overset{\kappa}{\longrightarrow} & B([\mu]) \\ & I \quad \searrow & \downarrow V \\ & & \mathbb{R} \end{array}$$

To prove the theorem, we are going to show that the diagram below also commutes

$$\begin{array}{ccccc} & & B(G) & & \\ & \tilde{\kappa} \nearrow & & \searrow \pi & \\ B(S) & \overset{k}{\longrightarrow} & B([\mu]) & \overset{h}{\longleftrightarrow} & B(G/PR(\mu)) \\ & I \searrow & \downarrow V & \swarrow & \\ & & \mathbb{R} & & \end{array}$$

and, then, define $\tilde{V}$ by composing the mappings on the right hand side of the diagram.

To begin, define $\tilde{\kappa}$ as in the statement of the theorem. Then, just as in Theorem 2, it is immediate to verify that $range(\tilde{\kappa}) \subset B(G)$.

Next, let $\theta \subset G$ denote a generic element of the partition of G produced by the action of $PR(\mu)$. We use the same notation to index elements in the quotient $G/PR(\mu)$. No confusion can result. From Lemma 3, $G/PR(\mu)$ is standard. Hence, there exists a canonical system of condition probabilities $\{P_\theta\}_{\theta \in G/PR(\mu)}$. For each $\psi \in B(G)$, define a the function $\pi(\psi) : G/PR(\mu) \to \mathbb{R}$ by

$$\pi(\psi)(\theta) = \int_\theta \psi \mid_\theta dP_\theta$$

It follows at once from Rokhlin's definition (Definition 2) that $\pi(\psi) \in B(G/PR(\mu))$. In other words, $\pi : B(G) \to B(G/PR(\mu))$.

Finally, recall that as we observed after Lemma 1, there exists a bijection between $G/PR(\mu)$ and $[\mu]$. By the preceding, $G/PR(\mu)$ is Polish and so is $[\mu]$ (considered as a subspace of $\mathcal{P}(S)$). Hence, by the Borel Isomorphism Theorem (**A.1**), they are Borel isomorphic. That is, there exists a Borel bijiection $b : G/PR(\mu) \to [\mu]$. Hence, it is immediate to see that if $f \in B([\mu])$, the mapping h defined by $f \longmapsto f \circ b$ is a bijection $B([\mu]) \to B(G/PR(\mu))$. □

B.6 Theorem 6

Proof of Theorem 6. In A.4, we observed that G embeds isometrically into the group of unitary operators, U, on $\mathcal{L}^2(\mu)$. Restricted to positive linear isometries, our equivalence relation $\sim$ is then the orbit equivalence relation produced by the action by conjugation of U on itself. Hence, the result follows from the fact that such an action is properly generically ergodic (see proof of Theorem 4). □

B.7 Proof of Proposition 1

Proof of Proposition 1. As noted in the proof of Theorem 3, any system $\{P_\iota\}_{\iota\in\mathcal{I}}$ of conditional probabilities (canonical or not) defines an operator $\tilde{T} : B(M) \longrightarrow \mathbb{R}^{M/\mathcal{I}}$ by

$$\psi \mapsto \tilde{T}(\psi) \qquad where \qquad \tilde{T}(\psi)(\iota) = \int\limits_M \psi dP_\iota$$

and the set

$$\Theta = \left\{ \psi \in B(M) \mid (a)\ \tilde{T}(\psi) \in B(M/\mathcal{I});\ (b) \int\limits_M \psi dP = \int\limits_{M/\mathcal{I}} \int\limits_{\iota} \psi \mid_\iota dP_\iota dP' \right\}$$

is a linear subspace. Since an act $f \in B(S)$ is subjectively measurable if and only if $\kappa(f) \in \Theta$, it follows that the class of subjectively measurable acts is the set $\kappa^{-1}(\kappa(B(S)) \cap \Theta)$. It is easily seen that this is a linear subspace of $B(S)$. Notice that $(\kappa(B(S)) \cap \Theta)$ is always non-empty because it contains the constant functions. From this, it follows that the class ΣME of subjectively measurable events contains $\varnothing$ and S and is closed under finite disjoint unions. Moreover, normality of κ implies that ΣME is closed under countable disjoint unions.

Finally, if an event $E \in \Sigma$ is unambiguous in the sense of [13] and [**?**], then $\kappa(\chi_E)$ is a constant mapping on M, that is $\chi_E \in \kappa^{-1}(\kappa(B(S)) \cap \Theta)$.

Finally, $\chi_E \in \Sigma ME$ implies $\int\limits_M \kappa(\chi_E) dP = \int\limits_{M/\mathcal{I}} \int\limits_\iota \kappa(\chi_E) \mid_\iota dP_\iota dP'$. Hence, $E \mapsto \int\limits_M \kappa(\chi_E) dP$ is a set function on ΣME which is countably additive. □

Columbia University, NY