TOWARD A RATIONAL-CHOICE FOUNDATION OF NON-ADDITIVE THEORIES

MASSIMILIANO AMARANTE

Université de Montréal et CIREQ

Abstract. A classical argument of de Finetti holds that Rationality implies Subjective Expected Utility (SEU). In contrast, the Knightian distinction between Risk and Ambiguity suggests that a rational decision maker would obey the SEU paradigm when the information available is in some sense good, and would depart from it when the information available is not good. Unlike de Finetti's, however, this view does not rely on a formal argument. In this paper, we study the set of all information structures that might be availabe to a decision maker, and show that they are of two types: those compatible with SEU theory and those for which SEU theory must fail. We also show that the former correspond to "good" information, while the latter correspond to information that is not good. Thus, our results provide a formalization of the distinction between Risk and Ambiguity. As a consequence of our main theorem (Theorem 2, Section 8), behavior not-conforming to SEU theory is bound to emerge in the presence of Ambiguity. We give two examples of situations of Ambiguity. One concerns the uncertainty on the class of measure zero events, the other is a variation on Ellberg's three-color urn experiment. We also briefly link our results to two other strands of literature: the study of ambiguous events and the problem of unforeseen contingencies. We conclude the paper by re-considering de Finetti's argument in light of our findings.

1. Introduction

In a problem of decision making under uncertainty, a decision maker has to rank a set of alternatives $\mathcal{A}$. Following Savage [17], these are modeled as mappings $(S, \Sigma) \longrightarrow X$, where (S, Σ) is a measurable space of states of the world and X

Key words and phrases. Rationality, Expected Utility, Non-additive theories, Information, Ambiguity

JEL classification: D81

I benefited from comments and suggestions from Tzachi Gilboa, Fabio Maccheroni, Massimo Marinacci, Rich McLean and Bill Zame.

is a space of consequences. Many models of decision making such as Subjective Expected Utility (SEU), Choquet Expected Utility (CEU) or Maxmin Expected Utility (MEU) focus on decision makers whose ranking $\succsim$ of the alternatives can be represented by means of a real-valued functional $I : \mathcal{A} \longrightarrow \mathbb{R}$, that is if f, $g \in \mathcal{A}$ then

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

The present paper aims at addressing the following question. Suppose that the decision maker's ranking is represented by a functional $I : \mathcal{A} \longrightarrow \mathbb{R}$; What properties does I have to have for the decision maker to be considered "rational"? Two seemingly incompatible answers are well-known. On the one hand, a classical argument of de Finetti holds that Rationality implies SEU. On the other hand, the Knightian distinction between Risk and Uncertainty (Ambiguity, in the current terminology) suggests that a rational decision maker would obey the SEU paradigm when the information available is in some sense good, and would depart from it when the information available is not good. While the first viewpoint relies on a formal argument (see Section 2 in this paper), it is not so for the second. Here, after reviewing the main ideas underlying the Knightian distinction as well as the fundamental structure of a wide class of models of decision making (Sections 3 to 6), we study (Section 7)) the set of all information structures that might be availabe to a decision maker. Our main result (Theorem 2, Section 8) unveils that information structures are of two types: those compatible with SEU theory and those for which SEU theory must fail. The former describe situations where the decision maker has the ability of distinguishing across all the scenarios that are relevant to his decision problem, while the latter describe situations where the decision maker lacks the ability to do so (Section 9). Transparently, one can interpret them as corresponding to "good" and "not good" information, respectively. These considerations come together into a formal definition of Knightian Uncertainty/Ambiguity, which we give in Section 10. By virtue of Theorem 2, behavior not-conforming to SEU theory is bound to emerge in the presence of Ambiguity. In Sections 11 and 12, we give two examples of situations of Ambiguity. One concerns the uncertainty on the class of measure zero events, the other is a variation on Ellberg's three-color urn experiment. Section 14 briefly links the results of this paper to two other strands of literature: the study of ambiguous events and the problem of unforeseen contingencies. Section

15 concludes by briefly re-considering de Finetti's argument. Four appendices supplement the material in the main text.

2. Rationality and Expected Utility

Following the seminal papers of Schmeidler [18] and of Gilboa and Schmeidler [11], during the past twenty years decision theorists have focused almost exclusively on non-additive theories, by which I mean theories that depart from the Expected Utility paradigm. For the most part, this has been motivated by the overwhelming experimental evidence indicating that, in many circumstances, actual people systematically violate the axioms of expected utility, in particular the independence axiom. Yet, usually no attempt is made to reconcile non-additive theories with some preconceived idea of rationality. This state of affairs has created a disparity between the status of non-additive theories and that of SEU theory, as the latter seems to be tightly related to a very appealing concept of rationality. This relation is expressed by de Finetti's famous no-arbitrage argument, which we now review.

Assume that the consequences associated with the various alternatives are expressed directly in terms of "utils". Hence, we can view alternatives as real-valued functions and, in fact, we may assume that the set of alternatives is the set $B(\Sigma)$ of bounded, Σ-measurable functions $(S, \Sigma) \longrightarrow \mathbb{R}$. Thus, an alternative $f \in B(\Sigma)$ can be thought of as a stock that pays off $f(s)$ utils if state $s \in S$ obtains. The starting point of de Finetti's argument is the idea that if a decision maker is to be considered rational, then he must not be subject to arbitrage. This seems to be a very mild requirement; in fact, it would be hard to support a definition of rationality that does not impose it. Now, suppose that the decision maker's ranking $\succsim$ is represented by a functional $I : B(\Sigma) \longrightarrow \mathbb{R}$, and suppose that I is not additive. Then, by definition, there exist $f, g \in B(\Sigma)$ such that $I(f+g) \neq I(f) + I(g)$. Without loss, assume that $I(f+g) > I(f) + I(g)$. Then, there exist prices p_{f+g}, p_f and p_g such that

$$I(f+g) > p_{f+g} > p_f + p_g > I(f) + I(g)$$

At these prices, the decision maker is willing to buy the stock $f+g$ and willing to sell the stocks f and g. But, by making these trades with the decision maker, one would be certain to make a profit, no matter which state of the world obtains.

Thus, we see that if the decision maker must not be subject to arbitrage, then his preference functional $I : B(\Sigma) \longrightarrow \mathbb{R}$ must be additive. In a similar fashion, the demand that the decision maker be immune to arbitrage leads to two more properties of the functional I. First, I must be monotone, that is if $f, g \in B(\Sigma)$ are such that $f(s) \geq g(s)$ for every s in S, then it must be the case that $I(f) \geq I(g)$. Second, by letting $\mathbf{1}$ denote the function identically equal to 1 on S, then it must be the case that $I(\mathbf{1}) = 1$. Summing up: if a decision maker must be immune to arbitrage, then his preference functional $I : B(\Sigma) \longrightarrow \mathbb{R}$ must be additive, monotone and such that $I(\mathbf{1}) = 1$. It is easy to see that these three properties together imply that I must be linear and continuous when $B(\Sigma)$ is equipped with the sup-norm topology (see Appendix A). Then, the Riesz representation theorem along with monotonicity and $I(\mathbf{1}) = 1$ imply that there exist a unique (finitely additive) probability measure P on Σ such that for all $f \in B(\Sigma)$

$$I(f) = \int_S f dP$$

We thus reach the remarkable conclusion that Rationality $\Longrightarrow$ SEU.

3. Why non-additive theories?

The conclusion reached in the previous section begs the obvious question: Why focus on non-additive theories? Two answers of a different nature are usually encountered. The first distinguishes between the normative and the positive aspect of a theory. The idea is that if one is concerned with the normative aspect only, then by virtue of de Finetti's argument only SEU theory should be considered. Things would be different, however, if one is concerned with explaining actual people's behavior. In this respect, a vast experimental evidence not only shows that actual people often systematically depart from the SEU paradigm, but also that there are discernible patterns in people's behavior. Hence, the interest for non-additive theories. I find this answer dissatisfactory for the following reason. Once we assume the validity of de Finetti's argument, we are forced to conclude that people who depart from the SEU paradigm are subject to arbitrage. Thus, these people are going to suffer losses, at least if there exists somebody capable of exploiting their weaknesses. It is then to be believed that over time these people would learn from their mistakes, modify their behavior and eventually approach the SEU paradigm: their departure from SEU should appear only as a temporary accident.

The second answer goes back to the idea of Knightian Uncertainty. Here, one takes steps from the observation that some of the implications of SEU theory are too strong to be intuitively sound. For one thing, SEU implies that the decision maker is always able (implicitly or explicitly) to assign probabilities to all events in Σ. Yet, in many circumstances the information available does not allow decision makers to do so. Thus, sound intuition suggests that precisely in those circumstances one should observe a departure from SEU. Notice that this view has two important implications. First, it must be the case that de Finetti's argument, if valid, relies implicitly on some assumption implying that the information available to the decision maker is "good enough". Second, a sound theory of decision making under uncertainty must account for the fact that the decision maker's behavior may vary substantially with the information available to him. In particular, the idea of Knightian Uncertainty suggests that the decision maker's behavior must obey additivity when the information is "good enough" and exhibit non-additivity when the information is not good enough. The remainder of this paper is devoted to explore whether or not this view is tenable.

4. Information and non-additivity

In the previous section, we identified a possible reason for studying non-additive theories: non-additivity might emerge in those circumstances where the information available to the decision maker is not good enough. This hypothesis suggests that, as a first step, we would look at situations where the information available is (in some sense) not good, and compare them with similar situations where the information available is good. This is the essence of Ellsberg's experiments as well as of an experiment devised by Gardenfors and Sahlin [7]. They consider the problem of betting on a tennis match, and focus on two situations.[1] In Situation 1, the decision maker is told that the two tennis players, player 1 and player 2, are equally good. In contrast, in Situation 2 the decision maker is told that one player is great and that the other is bad, but he is not told which one is which. We see that the decision maker is explicitly given information, and there is no

[1] They actually consider three different situations, but we only focus on two of those.

doubt that this information is qualitatively different across the two situations. The similarity with Ellsberg's experiments is obvious, and the same considerations that we are going to make below could be made for Ellsberg's experiments as well.

In Situation 1, the information given to the decision maker transparently translates into the fact that one must assign a value of $1/2$ to the event that player 1 wins the match. What about Situation 2? If one believes de Finetti's argument, then the rational decision maker must have a probability. Since the information is symmetric in the labels 1 and 2 assigned to the players, this probability must be symmetric in those labels. We then conclude that the decision maker must assign a value of $1/2$ to the event that player 1 wins the match. Thus, despite the obvious qualitative difference between Situation 1 and Situation 2, a de Finetti-rational decision maker must exhibit the same behavior in both situations: since the probability governing the decision maker's behavior is the same, in Situation 2 the decision maker must take exactly the same set of bets that he would take in Situation 1. Not surprisingly, this conclusion is not supported by the observation of actual people's behavior. In fact, in Situation 2 many people choose to not bet at all (see [7]), a behavior that *per se* implies a violation of the SEU paradigm.

Another implication of the view that Rationality $\Longrightarrow$ SEU is as follows. Suppose that in each situation a large number of matches takes place, but the outcome of each match is not observed by the decision maker. Suppose now that the decision maker is asked to provide a point estimate e of the ratio $x =$ (victories of 1/ total number of matches). Equivalently, consider (from the viewpoint of the decision maker) the maximization problem

$$\max_{e} - |e - x|$$

It is easy to see that the value of the problem converges to 0 (as the number of matches goes to infinity) in Situation 1, while it is always $-(1/2)$ in Situation 2. In particular, $1/2$ is a best point-estimate for x in both circumstances, even though in Situation 2 the decision maker is certain that this estimate is wrong. Thus, the view that a rational decision maker must have a probability leads to the bizzarre outcome of a best point-estimate which is known to be wrong. These observations support, and perhaps strengthen, the point of the previous section: in certain circumstances, not only can one not assign probabilities to all relevant events but also to insist that one would do so leads to behavior that can hardly be considered rational.

5. Is non-additivity compatible with rationality?

In the previous sections, we reviewed de Finetti's argument that Rationality implies SEU; we then reviewed the idea of Knightian Uncertainty; finally, we saw that when the information is of a certain type, de Finetti's idea leads to questionable conclusions regarding the decision maker's rationality. As stated in Section 3, the remainder of the paper will be mainly devoted to exploring the idea that non-additive behavior might emerge as a rational response to those situations characterized by the fact that the information available is (in some sense) not good. Our strategy will be as follows. First, we want to understand what is the formal courterpart of statements like "the information is good" or "the information is not good enough". This is a necessary and crucial step. The formalization of the Knightian distinction between Risk and Ambiguity demands that we would be able to distinguish among situations according to the quality of the information available. Once this is accomplished, we are going to consider the problem of whether or not it is possible to obey the SEU paradigm when the information available is not good enough. We will find (Theorem 2, Section 8) that the answer is negative, thus providing the grounds for the emergence of non-additive behavior in the presence of information which is "not good". We will then re-consider examples *à la* Ellsberg as well as other types of examples. Finally, we will re-consider de Finetti's argument in light of our findings.

Notation: Throughout the paper, we will be using the following notation. The set of bounded, Σ-measurable functions $(S, \Sigma) \longrightarrow \mathbb{R}$ equipped with the sup-norm is denoted by $B(\Sigma)$. Its dual $ba(\Sigma)$, the space of bounded charges on Σ, is always endowed with the weak*-topology produced by the duality $(ba(\Sigma), B(\Sigma))$. The subset of $ba(\Sigma)$ consisting of the finitely additive probability measures on Σ is denoted by $ba_1^+(\Sigma)$. For $\mathcal{C}$ a weak*-compact, convex subset of $ba_1^+(\Sigma)$, a weak*-continuous affine function $\mathcal{C} \longrightarrow \mathbb{R}$ is of the form $\psi_f(P) = \int\limits_S f dp$, $P \in \mathcal{C}$, for some $f \in B(\Sigma)$. The space of all weak*-continuous affine functions on $\mathcal{C}$ equipped with the sup-norm is denoted by $A(\mathcal{C})$. The mapping $\kappa : f \longmapsto \psi_f$ is the canonical linear mapping $\kappa : B(\Sigma) \longrightarrow A(\mathcal{C})$. The Borel σ-algebra on $\mathcal{C}$ is denoted by $\mathcal{B}$, and $B(\mathcal{B})$ denotes the space of bounded, $\mathcal{B}$-measurable functions $\mathcal{C} \longrightarrow \mathbb{R}$ equipped with the sup-norm. Finally, the set of regular Borel measures on $\mathcal{C}$ is denoted by $\mathcal{P}(\mathcal{C})$.

6. Additive vs non-additive theories: the point of departure

In [9], Ghirardato, Maccheroni and Marinacci isolated a core common to several theories of decision making. This consists of the five axioms listed below. Let $\mathcal{A}$ denote the set of all alternatives, and let $\mathcal{A}_c$ be that of constant alternatives, that is of constant mappings $(S, \Sigma) \longrightarrow X$. Assume that X is a mixture space (see [3] and [10]), and let $\succsim$ denote the decision maker's preference relation over $\mathcal{A}$.

A1 $\succsim$ is complete and transitive.

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

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

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

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

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

Then, Ghirardato, Maccheroni and Marinacci observed that alternative sixth axioms correspond to alternative theories of decision making. For instance, one obtains SEU, CEU and MEU as follows:

A6 (a) (SEU, Anscombe and Aumann [3]) For all $f, g \in \mathcal{A}$ such that $f \sim g$, $\frac{1}{2}f + \frac{1}{2}g \sim f$;

A6 (b) (CEU, Schmeidler [18]) For all $f, g \in \mathcal{A}$ such that $f \sim g$, $\frac{1}{2}f + \frac{1}{2}g \sim f$ if f and g are comonotonic;

A6 (c) (MEU, Gilboa and Schmeidler [11]) For all $f, g \in \mathcal{A}$ such that $f \sim g$, $\frac{1}{2}f + \frac{1}{2}f \succsim f$.

As discussed above, we want to explore the idea that non-additive behavior might emerge when the information available is in some sense not good. To do so, we must focus on the point where SEU and non-additive theories depart from each other. By virtue of the Ghirardato-Maccheroni-Marinacci's result, this demands that we would proceed as follows. First, we must fully characterize the structure implied by the first five axioms. Once this is done, we can then ask what conditions have to be met for a certain sixth axiom to be valid. In particular, we can ask what is the informational content (if any) embedded in the axiom that delivers SEU.

The first step was accomplished in [2, Theorems 1 and 2], which yielded the characterization below. Recall that axioms 1 to 5 imply that there exist a utility function $u : X \longrightarrow \mathbb{R}$ and a functional $I : B(\Sigma) \longrightarrow \mathbb{R}$ such that for $\tilde{f}, \tilde{g} \in \mathcal{A}$ (see [11] and [9], for details)

$$\tilde{f} \succsim \tilde{g} \qquad iff \qquad I(u \circ \tilde{f}) \geq I(u \circ \tilde{g})$$

For notational simplicity, throughout the paper we are going to indentify an act $\tilde{f} \in \mathcal{A}$ with the corresponding function $u \circ \tilde{f} = f \in B(\Sigma)$.

Theorem 1 (Amarante [2])**.** *A preference relation $\succsim$ on $\mathcal{A}$ satisfies axioms 1 to 5 iff*

(a) The functional I representing it factors as $I = V \circ \kappa$, where κ is the canonical linear mapping $B(\Sigma) \longrightarrow A(\mathcal{C})$, $\mathcal{C}$ a weak-compact, convex subset of $ba_1^+(\Sigma)$ and $V : A(\mathcal{C}) \longrightarrow \mathbb{R}$ is monotone and comonotonic additive (hence, sup-norm continuous).*

In turn, condition (a) is equivalent to the following

(b)There exists a capacity ν on the Borel subsets of $\mathcal{C}$ such that for any $f \in B(\Sigma)$

$$I(f) = \int_{\mathcal{C}} \kappa(f) d\nu$$

Intuitively, Theorem 1 tells us that any behavior satisfying axioms 1 to 5 corresponds to (is represented by) an integration over priors when this operation is performed in the sense of Choquet. It is now (deceivingly) easy to see what it takes for SEU theory to obtain.

Corollary 1. *A preference relation $\succsim$ on $\mathcal{A}$ is a SEU preference iff the capacity in Theorem 1 is a measure (i.e., $\nu \in \mathcal{P}(\mathcal{C})$).*

Proof. If $\nu \in \mathcal{P}(\mathcal{C})$, then by [15, Proposition 1.1] ν has a unique barycenter $P^* \in \mathcal{C}$, that is P^* is such that $\int_{\mathcal{C}} \kappa(f) d\nu = \kappa(f)(P^*) = \int_S f dP^*$ for every $f \in B(\Sigma)$; thus SEU obtains. Conversely, if the preference is SEU, then the functional I is linear. By Theorem 1 part (a), since κ is linear and sup-norm continuous, we conclude that V must be linear and sup-norm continuous. By Hahn-Banach, V can be extended to a continuous linear functional on $C(\mathcal{C})$, the Banach space of all continuous functions on $\mathcal{C}$ equipped with the sup-norm, and (via the Riesz representation theorem) there exists a unique regular Borel measure representing it. That is, $\nu \in \mathcal{P}(\mathcal{C})$. □

It remains to be seen what is the informational content (if any) embedded in the requirement that the capacity ν be a measure (equivalently, that V be linear and continuous). We will study this in the next section.

7. Information

We now want to take a closer look at Corollary 1, and see what it really shows. The Corollary says that a SEU decision maker can be thought of as somebody who (Lebesgue) integrates over priors. Formally, in the terminology of Theorem 1, this corresponds to the ability of integrating all the functions in $\kappa(B(\Sigma))$. Thus, whatever we mean by "information", we must conclude that what Corollary 1 really tells us is that *the SEU decision maker has enough information to integrate all the functions in* $\kappa(B(\Sigma))$. Equivalently, suppose that we have formalized the concept of information, that is we have associated the intuitive idea of information to a class of mathematical objects. Then, Corollary 1 says that the mathematical object corresponding to the decision maker in Corollary 1 contains (an equivalent of) the class of Borel subets of $\mathcal{C}$. This is so because knowledge of this class is necessary to be able to integrate all the functions in $\kappa(B(\Sigma))$.

These considerations suggest that we should look deeper into the concept of information and into its formalization. The reader should notice that we have been speaking of information about the set of measures $\mathcal{C}$. In passing, we observe that this is the type of information that decision makers are given in the Ellsberg's experiments (the configurations of the urns), in those of Gardenfors and Sahlin (the ability of the players) and, to our knowledge, in all those circumstances associated with systematic departures from SEU. At any rate, information of the form "the true state belongs to the set $A \subset S$" can always be trivially expressed as information about the set $\mathcal{C}$.

7.1. **What is information?** In its most basic aspect, "information" has to do with the ability of distinguishing among things and, ultimately, this boils down to the ability (or lack thereof) of saying that A has a certain property that B does not have. It is easy to translate this idea into a basic mathematical concept. Suppose that we are interested in a certain collection of things, like the collection $\mathcal{C}$ of measures on (S, Σ) encountered above. By definition, any subset of $\mathcal{C}$ corresponds to a certain property: it is the collection of all those elements of $\mathcal{C}$ that have that property. Given two points P and Q in $\mathcal{C}$, a decision maker can distinguish between P and Q if he can point at a property that P has and Q does not have, and cannot distinguish between P and Q if he does not know of one such property. Equivalently, he can distinguish between P and Q if he can point at a subset of $\mathcal{C}$ which contains P but does not contain Q. Thus, the information that a decision maker has about a certain set E is described by the list of all the properties he has knowldege of, which, in turn, corresponds to a certain collection of subsets of E. Formally,

Definition 1. *Let E be a set. An Information about E is a pair $(E, \mathcal{E})$, where $\mathcal{E}$ is a collection of subsets of E.*

7.2. **The informational content of Corollary 1.** We can now formally describe the information available to the decision maker of Corollary 1.

Corollary 2. *A preference relation $\succsim$ on $\mathcal{A}$ is a SEU preference iff*

(a) The capacity in Theorem 1 is a measure; and

(b) The decision maker's information about $\mathcal{C}$ contains the Borel field generated by the weak-topology on $\mathcal{C}$.*

As we discussed above, part (b) of Corollary 2 describes explicitly the condition – implicit in Corollary 1 – that the decision maker is able to integrate all the functions in $\kappa(B(\Sigma))$.

We have been wanting to explore the hypothesis that a rational decision maker obeys the SEU paradigm when the information available to him is good, and departs from it when the information is not good. So the result in Corollary 2 begs the obvious question: Is the information described by the Borel field good or not good? The Borel field generated by the weak*-topology on $\mathcal{C}$ separates points. This means that *for any* two points P and Q in $\mathcal{C}$, the decision maker always knows of a property that P has and Q does not have. Clearly, this is very good information as things cannot get better than this. For a more concrete depiction of the situation, the reader should think of a decision maker who is going to observe the realizations associated with a certain phenomenon, and who wants to test the hypothesis that the sample is drawn from process P rather than process Q, with $\mathcal{C}$ being the set of all possible processes. To say that the decision maker's information about $\mathcal{C}$ separates points in $\mathcal{C}$ concretely means that, for any two points P and Q in $\mathcal{C}$, (a) the decision maker is able to test the hypothesis that the true process is P rather than Q; and (b) he knows that there exists an observable sample for which the hypothesis is either accepted or rejected.

Summing up, the result of Corollary 2 is consistent with the hypothesis that a rational decision maker obeys SEU theory when the information available to him is good. Yet, Corollary 2 does not say anything about what happens when the information available is not as good as that described by the Borel field.

7.3. **Coarse information structures.** We now turn our attention to those situations where the information available about $\mathcal{C}$ is not as good as that described by the Borel field $\mathcal{B}$ generated by the weak*-topology on $\mathcal{C}$. What we have to do is to study how the decision maker's behavior varies when his information is a proper sub-collection of $\mathcal{B}$. We will limit our attention to sub-collections that are sub-algebras of $\mathcal{B}$; in fact, sub-algebras generated by partitions of $\mathcal{C}$. Consideration of more general collections is certainly possible by virtue of Definition 1, but we do not believe that that would add to the conceptualization that we will provide in this paper.

Definition 2. *An information structure on* $(\mathcal{C},\mathcal{B})$ *is a triple* $\{(\mathcal{C},\mathcal{B}),\mathcal{I},\mathcal{B}_{\mathcal{I}}\}$*, where* $\mathcal{I}$ *is a partition of* $\mathcal{C}$ *and* $\mathcal{B}_{\mathcal{I}}$ *is the sub-field of* $\mathcal{B}$ *generated by* $\mathcal{I}$*.*

The partition and the associated sub σ-field state that the decision maker has only partial information about $\mathcal{C}$. In terms of the example discussed above, this corresponds to the following situation (see Billingsley [4], 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 process lies. However, such a decision maker would not be able, *on the basis of his information*, to construct an experiment capable of distinguishing among processes lying in the same cell of the partition.

7.4. **Informational constraints.** The partition of the decision maker of Corollary 2 is the one generated by the identity mapping on $\mathcal{C}$: the decision maker can distinguish between any two points in $\mathcal{C}$. When his partition is coarser, his ability of distinguishing between points is limited as is his ability of constructing statistical experiments. The formal counterpart of these limitations consists of a certain amount of restrictions that the functional $V : A(\mathcal{C}) \longrightarrow \mathbb{R}$ in Theorem 1 must satisfy. The remainder of this section is devoted to describing these constraints. Our exposition is going to be a bit pedantic. We feel, however, that this, if not strictly necessary, is at least justified by the developments of the next sections.

Given a partition $\mathcal{I}$ of $\mathcal{C}$, the functional $V : A(\mathcal{C}) \longrightarrow \mathbb{R}$ must respect two types of constraints. First, the evaluation of a function $\varphi \in A(\mathcal{C})$ must respect the information available; second, only those functions in $A(\mathcal{C})$ for which "enough information" is available can be evaluated. Limitations of the first type express the fact that all that the decision maker can get to know is an element ι of the partition. For $\varphi \in A(\mathcal{C})$, let us denote by $V(\varphi \mid \iota)$ the evaluation of φ given that ι obtains. Then, we must have

$$\varphi, \psi \in A(\mathcal{C}) \text{ and } V(\varphi \mid \iota) = V(\psi \mid \iota) \text{ for all } \iota \in \mathcal{I} \implies V(\varphi) = V(\psi) \tag{7.1}$$

That is, if two functions in $A(\mathcal{C})$ 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 (7.1) can be restated in a more useful way. Let $\mathcal{C}/\mathcal{I}$ denote the quotient of $\mathcal{C}$ by the partition $\mathcal{I}$, that is the space whose elements are the cells of the partition. Define a mapping $\tilde{\pi}_V : A(\mathcal{C}) \longrightarrow \mathbb{R}^{\mathcal{C}/\mathcal{I}}$ by $\varphi \longmapsto (V(\varphi \mid \iota))_{\iota \in \mathcal{I}}$; that is, each function $\varphi \in A(\mathcal{C})$ is associated to the real function defined on the quotient $\mathcal{C}/\mathcal{I}$ that at point ι (viewed as point in the quotient) takes value $V(\varphi \mid \iota)$. Then, condition (7.1) says that the functional $V : A(\mathcal{C}) \longrightarrow \mathbb{R}$ must be expressible by means of the diagram below[2]

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

[2]Of course, whenever V satisfies condition (7.1), then there exists a unique functional V' which makes the diagram commute.

The second type of constraint is that the decision maker can evaluate (measurably) only those functions that are measurable with respect to his information, which is represented by $\mathcal{B}_{\mathcal{I}}$, the σ-field generated by the partition. This is equivalent to the requirement that the domain of the functional V' above must consists of functions $\mathcal{C}/\mathcal{I} \longrightarrow \mathbb{R}$ which are measurable with respect to the canonical σ-field on $\mathcal{C}/\mathcal{I}$.[3] This leads to the following observation: *a function $\varphi \in A(\mathcal{C})$ can be measurably evaluated if and only if $\tilde{\pi}_V(\varphi)$ is a measurable function $\mathcal{C}/I \longrightarrow R$.*

Summing up, the decision maker can (measurably) evaluate all the functions in $A(\mathcal{C})$ *while respecting his information* if and only if two conditions are satisfied:

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

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

7.5. **Bayesian information structures.** A necessary condition for a decision maker to be Bayesian is that his functional V in Theorem 1 is linear and (sup-norm) continuous. Thus, V can be represented by a measure, a *prior*, on $\mathcal{B}$. For these decision makers, the information available is described not only by the partition but also by the prior. For instance, if the prior is concentrated on a single point $P \in \mathcal{C}$, then the decision maker is certain that P is the true scenario independently of the restrictions imposed by the partition. Clearly, this is a situation where the information available to the decision maker is very good: For any $Q \in \mathcal{C}$, $Q \neq P$, the decision maker can test the hypothesis that the true measure is Q, and will reject it with probability 0 of making an error. Thus, as the example makes it clear, whenever we want to assess the quality of information available to this type of decision makers, we would have to refer jointly to the prior and to the partition. This motivates the following definition

[3]This is the finest σ-field on $\mathcal{C}/\mathcal{I}$ for which the canonical projection $(\mathcal{C}, \mathcal{B}) \longrightarrow \mathcal{C}/\mathcal{I}$ is measurable. It is easy to see that the σ-field on $\mathcal{C}/\mathcal{I}$ must be such that the canonical projection is measurable. For, if not, we would reach the absurd conclusion that the decision maker has more information than the one described by $\mathcal{B}$. To see this, suppose, by the way of contradiction, that the decision maker has a field on $\mathcal{C}/\mathcal{I}$ for which the canonical projection is not measurable. Thus, there exists an event A in $\mathcal{C}/\mathcal{I}$ for which $\pi^{-1}(A) \notin \mathcal{B}$. By knowing the set $\mathcal{C}$ and his field on $\mathcal{C}/\mathcal{I}$, the decision maker has enough information to evaluate the bet $\chi_{\pi^{-1}(A)}$ ($\chi_{\cdot}$ denotes indicator functions): 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.

Definition 3. *A Bayesian information structure on $(\mathcal{C}, \mathcal{B})$ is a quadruple $\{(\mathcal{C}, \mathcal{B}), \mu, \mathcal{I}, \mathcal{B}_{\mathcal{I}}\}$, where μ is a regular Borel measure on $(\mathcal{C}, \mathcal{B})$, $\mathcal{I}$ is a partition of $\mathcal{C}$ and $\mathcal{B}_{\mathcal{I}}$ is the sub-field of $\mathcal{B}$ generated by $\mathcal{I}$.*

Remark 1. *The requirement that the measure μ in the definition be a regular Borel measure is motivated by the proof of Corollary 1. Consideration of wider classes of measures, in particular of finitely additive measures is possible. This would only strengthen our results, as it is shown in Section 13.*

When V is represented by a measure, the mapping $\tilde{\pi}_V$ in the diagram in Section 7 is the familiar conditional expectation operator, and the requirements in 7.4 correspond to the existence of a canonical system of conditional measures (see Appendix C). Intuitively, the idea that SEU would obtain goes as follows. Given his prior μ on $\mathcal{C}$ and the informational constraints expressed by the partition $\mathcal{I}$, the decision maker computes a collection of conditional probabilities, one for each element of the partition. Then, he averages these conditionals with the weights that μ gives to the corresponding elements of the partition, and SEU obtains. In the next section, however, we are going to see that this is not guarateed to work.

8. The impossibility of being Bayesian

With Corollary 2 and the discussion following it, we have seen that it is possible to conform to the SEU paradigm when the information available is very good, in fact the best possible. In this section, we want to see whether or not one can always conform to the SEU paradigm, independently of the information available. We are going to make the following assumption on the set of measures $\mathcal{C}$ of Theorem 1

Assumption: $\mathcal{C}$ is a Polish space.

The assumption is always satisfied whenever $\mathcal{C}$ is finite dimensional. More generally, in Appendix D we show that the assumption is satisfied whenever (S, Σ) is a Standard Borel space (see Appendix B, for a definition) and the decision maker's preference relation satisfies the axiom of Monotone Continuity (see Appendix D). The reason for introducing the assumption is technical: it allows us to use a result of Rokhlin [16] on the characterization of non-measurable partitions, whose definition we recall next. We stress that the concept refers to a property of the partition as a whole and not to a property of the sets making up the partition which, in this paper, are always assumed to be measurable.

Definition 4 (Rokhlin [16]). *Let (L, Λ, λ) be a Lebesgue space (see Appendix B), and let $\mathcal{I}$ be a partition of L. Let the quotient $L/\mathcal{I}$ be endowed with the measure structure induced by the canonical projection. The quotient $L/\mathcal{I}$ is said to be countably separated if there exists a countable family of measurable subsets of $L/\mathcal{I}$ which separates points. The partition $\mathcal{I}$ is called measurable if $L/\mathcal{I}$ is countably separated.*

As noted above, the assumption that the decision maker has a Bayesian information structure is necessary for the validity of SEU theory. The theorem below shows, however, that this condition is not sufficient. In fact, the theorem shows that SEU theory fails if (and only if) the information structure consists of both a prior with a non-atomic part and a non-measurable partition. We will comment on these features extensively in the next sections.

Theorem 2. *Let $\{(\mathcal{C}, \mathcal{B}), \mu, \mathcal{I}, \mathcal{B}_{\mathcal{I}}\}$ be a Bayesian information structure on $(\mathcal{C}, \mathcal{B})$. Assume that the prior μ is not purely atomic and that μ is not supported by a single cell of the partition. Then, SEU obtains if and only if the partition $\mathcal{I}$ of $\mathcal{C}$ is measurable.*

The assumption that μ is not supported by a single cell of the partition is clearly necessary for the conclusion in the theorem. In fact, if μ is supported by a single cell, then (modulo sets of μ-measure 0) the partition consists of a single element, the conditional measure coincides with the prior and SEU obtains.

Proof. By the assumption that μ is not purely atomic, μ can be expressed as the product of a purely atomic measure and a non-atomic one. Since a system of conditional measures of a purely atomic measure always exists, we can assume without loss that μ is non-atomic. If $\mathcal{I}$ is measurable, then by Rokhlin's Theorem [16] there exists a canonical system of conditional probabilities $\{\mu_\iota\}_{\iota \in \mathcal{I}}$. By using Definition 7 (Appendix C), it is straightforward to check that for every $\varphi \in B(\mathcal{B})$, we have

$$\int_{\mathcal{C}} \varphi d\mu = \int_{\mathcal{C}/\mathcal{I}} \int_{\iota} \varphi \mid_\iota d\mu_\iota d\mu'; \quad \iota \in \mathcal{I}, \quad \varphi \in B(\mathcal{B})$$

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

$$\int\limits_{\mathcal{C}}\varphi d\mu=V(\varphi)=\int\limits_{\mathcal{C}/\mathcal{I}}\int\limits_{\iota}\varphi\mid_{\iota} d\mu_{\iota}d\mu'=V'\circ T_{\mathcal{I}}(\varphi)\qquad,\qquad\forall\varphi\in B(\mathcal{B})$$

That is, both conditions (*) and (**) in Section 7.4 are satisfied. Finally, recall that the decision maker orders acts in $B(\Sigma)$ 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 the Riesz representation theorem.

Conversely, let $\mathcal{I}$ be nonmeasurable, and let $\{\mu_{\iota}\}_{\iota\in\mathcal{I}}$ be a system of conditional probabilities, with each μ_{ι} a non-atomic measure on ι. By Rokhlin's theorem, $\{\mu_{\iota}\}_{\iota\in\mathcal{I}}$ cannot be canonical. Hence, $\exists\varphi\in B(\mathcal{B})$ such that either $T_{\mathcal{I}}(\varphi):\mathcal{C}/\mathcal{I}\longrightarrow\mathbb{R}$ is nonmeasurable or $\int_{\mathcal{C}}\varphi d\mu\neq\int_{\mathcal{C}/\mathcal{I}}\int_{\iota}\varphi\mid_{\iota} d\mu_{\iota}d\mu'$. Either way, at least one of the conditions (*) and (**) of Section 7.4 is violated. If such a φ belongs to *range* $\kappa(B(\Sigma))$, then we are done. Now, we are going to show that *range* $\kappa(B(\Sigma))$ necessarily contains at least one such φ.

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

$$\psi\mapsto\tilde{T}(\psi)\qquad where\qquad\tilde{T}(\psi)(\iota)=\int\limits_{\mathcal{C}}\psi d\mu_{\iota}$$

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

$$\Theta=\left\{\psi\in B(\mathcal{B})\mid(a)\ \tilde{T}(\psi)\in B(\mathcal{B}/\mathcal{I});\ (b)\ \int\limits_{\mathcal{C}}\psi d\mu=\int\limits_{\mathcal{C}/\mathcal{I}}\int\limits_{\iota}\psi\mid_{\iota} d\mu_{\iota}d\mu'\right\}$$

B using standard arguments, it is easily checked (see for instance [1], Ch. 13) that Θ is a linear subspace and a lattice. Now, let $\{\psi_n\}_{n\in\mathbb{N}}$ be a sequence in Θ;

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

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

(a) By the Dominated Convergence Theorem (DCT), for every μ_{ι} we have $\int_{\mathcal{C}}\psi_n d\mu_{\iota}\nearrow\int_{\mathcal{C}}\psi d\mu_{\iota}$, 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 $\psi \in B(\mathcal{B})$, $\tilde{T}(\psi)$ is bounded, i.e. $\tilde{T}(\psi) \in B(\mathcal{B}/\mathcal{I})$.

(b) Observe that

$$\begin{aligned}
\int_{\mathcal{C}} \psi d\mu &= \lim_{n\to\infty} \int_{\mathcal{C}} \psi_n d\mu && \textit{(by the DCT and } \psi \in B(\mathcal{B})\textit{)} \\
&= \lim_{n\to\infty} \int_{\mathcal{C}/\mathcal{I}} \int_{\iota} \psi_n d\mu_\iota d\mu' && \text{(because } \psi_n \in \Theta\text{)} \\
&= \lim_{n\to\infty} \int_{\mathcal{C}/\mathcal{I}} \tilde{T}(\psi_n) d\mu' \\
&= \int_{\mathcal{C}/\mathcal{I}} \tilde{T}(\psi) d\mu' && \textit{(by (a) and the DCT)}
\end{aligned}$$

$$= \int_{\mathcal{C}/\mathcal{I}} \int_{\iota} \psi \quad | \quad {}_\iota d\mu_\iota d\mu'$$

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* $\kappa(B(\Sigma)) \subset \Theta$. Let K denote the set of continuous, convex functions on $\mathcal{C}$. Then, if $\gamma \in K$ there exists ([15], p. 19) $\{\zeta_m\}_{m\in\mathbb{N}} \subset A(\mathcal{C}) \subset$ *range* $\kappa(B(\Sigma))$ and a sequence $\{\eta_n\}_{n\in\mathbb{N}}$, with $\eta_n = \wedge \{\zeta_i\}_{i=1}^k$, such that $\eta_n \nearrow \gamma$. The sequence $\{\eta_n\} \subset \Theta$ because Θ is a lattice. Then, 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 $C(\mathcal{C})$, the set of continuous functions on $\mathcal{C}$. Since $\mathcal{C}$ is a metric space, for any closed set $A \subset \mathcal{C}$, there exists ([1], Corollary 3.14) $\{\lambda_n\}_{n\in\mathbb{N}} \subset C(\mathcal{C})$ such that $\lambda_n \searrow \chi_A$, where χ_A denotes the indicator function of A. Since $K - K$ is uniformly dense in $C(\mathcal{C})$, 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

$$\lambda_0(P) - 1 < h_{0_{k0}}(P) < \lambda_0(P) + 1 \qquad , \qquad \forall P \in \mathcal{C}$$

Then, the function

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

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

$$\lambda_0(P) + 1 < g_0(P) < \lambda_0(P) + 3 \qquad , \qquad \forall P \in \mathcal{C}$$

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

$$\lambda_1(P) - \frac{1}{3} < h_{1_{k1}}(P) < \lambda_1(P) + \frac{1}{3} \qquad , \qquad \forall P \in \mathcal{C}$$

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

$$\lambda_1(P) + \frac{1}{3} < g_1(P) < \lambda_1(P) + 1 \qquad \forall P \in \mathcal{C}$$

Moreover, for every $P \in \mathcal{C}$, we have

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

Inductively, define

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

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

$$\sup_{P\in\mathcal{C}} |g_n(P) - \lambda_n(P)| < \frac{1}{3^{n-1}}$$

Now, the inequality

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

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

By the above claim, we then have $\chi_A \in \Theta$ for any closed set $A \subset \mathcal{C}$. Next, observe that:

(i) $\chi_{\mathcal{C}} \in \Theta$ because the function $1 \in A(\mathcal{C}) \subset \Theta$;

(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 $\left\{\chi_{A_n}\right\} \subset \Theta$, then $\chi_{A_n} \nearrow \chi_A$ and $\chi_A \in \Theta$ by the claim above.

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

9. The meaning of the non-measurability of the partition

In the previous section, we have identified the class of information structures for which SEU theory fails. Two ingredients are necessary: the prior has to be diffused and the partition has to be non-measurable. Now, we want to understand whether or not these information structures are the ones, and the only ones, that correspond to the intuitive notion of information which is "not good enough".

The condition that the prior be diffused is quite transparent. A purely atomic prior, being concentrated on a countable number of points, is clearly associated to very good information: modulo sets of measure zero, the set $\mathcal{C}$ consists of countably many points and, by definition, the decision maker can distinguish between any two of them. Thus, this is a case where the information available is the best possibile. Correspondingly, the quotient space is trivially countably separated, the partition is measurable and SEU theory obtains. So, the condition that the prior have a non-atomic part is necessary if we want to talk of information which is, in some sense, not good. By Theorem 2, the issue now boils down to understanding the meaning of the condition that the partition be non-measurable. In this section, we are going to elucidate this by means of two examples. The goal is to highlight the following feature: a partition is non-measurable if the decision maker cannot distinguish, *on the basis of his information*, between (at least two) cells of the partition.

The examples we are going to describe are the most popular examples of non-measurable partitions. They 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 α is a fixed irrational number.

Topological properties. Example (1): Begin by considering 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. 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. Now, suppose that the decision maker's information about the torus consists of its topology, that is the class of open subsets of the torus. As we said, an open set corresponds to a certain property, and suppose that the decision maker tries to distinguish between any two spirals on the basis of that property. Since each spiral is dense in the torus, all the equivalence classes intersect that open set. Thus, the decision maker cannot distinguish among spirals on the basis of that property. But, the denseness of each spiral implies that the situation is the same for any other open set. Hence, the inability of distinguishing among spirals. The property that each equivalence class is dense (and meager) translates into a property of the quotient space: its topology (the finest topology which makes the canonical projection continuous) does not separate points. In fact, it is easy to see that the only closed sets in the quotient are the empty set and the whole quotient. Thus, while as a set the quotient has uncountably 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 $f : x \longmapsto x + \alpha \pmod 1$. 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 about the quotient as seen above obtains.

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

Measure space properties. 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. The problem of finding a countable separating family (Definition 4) for ξ is equivalent to the problem of finding a countable separating family for the partition η of the unit circle T defined as follows. Two points $x, y \in T$ are in the same cell of the partition η if and only if $\exists n \in \mathbb{N}$ such that $y = x + n\alpha \pmod 1$. Hence, if we define the map

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

we see that the elements of the partition are precisely the r_α-invariant subsets of T. It is well-known [5], 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.

10. A formalization of Knightian Uncertainty

As we saw above, what happens when the partition is non-measurable is that for each and every property that the decision maker knows of, each cell of the partition has at least a point that has that property. Equivalently, on the basis of his information the decision maker is unable to distinguish among the different cells of the partition. The intuition in terms of statistical experiments is similar: for each and every outcome the decision maker might observe, there is a process in each cell of the partition that in consistent with that outcome. In contrast, when the partition is measurable, there always exists a property that the decision maker knows of, and that allows him to distinguish between any two equivalence classes. Thus, the concept of non-measurable partition is the formal conterpart of the intuitive idea of insufficient information. We record this in the following definition.

Definition 5. *A decision maker faces Knightian Uncertainty (or Ambiguity) whenever his information about the set $\mathcal{C}$ is described by a quadruple $\{(\mathcal{C}, \mathcal{B}), \mu, \mathcal{I}, \mathcal{B}_{\mathcal{I}}\}$ (see Definition 3) with the following properties:*

(i) μ contains a non-atomic part;

(ii) μ is not concentrated on a single equivalence class;

(iii) the partition $\mathcal{I}$ (modulo μ-measure zero events) is non-measurable.

In correspondence to all other information structures, the decision maker faces (Knightian) Risk.

Theorem 2 can now be reformulated as follows: *If the decision maker faces Knightian Uncertainty, then he cannot obey the SEU paradigm.*

We have thus completed the program outlined in Section 5. We have partitioned the set of all (Bayesian) information structures into two subsets: one represents those information structures which describe good information, and corresponds to situations of Risk; the other represents those information structures which describe information which is not good, and corresponds to situations of Ambiguity. A Bayesian decision maker obeys SEU theory when he faces Risk, and departs from it when he faces Ambiguity.

The next question we have to ask is how substantial this is. That is, is our concept of Ambiguity representative of only some mathematical pathology with no bearing on actual decision makers, or rather is it descriptive of situations faced by actual decision makers? In the next sections, we are going to show, by means of examples, that the concept of Ambiguity in Definition 5 is far from being a mathematical curiosity.

11. Uncertainty on the measure zero class

In this and in the example of the next section, we are going to assume that the state space (S, Σ) is an uncountable Polish space. By the Borel Isomorphism Theorem (Appendix B), without loss we can think of (S, Σ) as of the interval $[0, 1]$ equipped with its usual Borel structure. For the purpose of interpretation, the assumption that S is uncountable deserves some explanation. Given a set $\mathcal{C}$ of measures on (S, Σ) and a partition $\mathcal{I}$ of $\mathcal{C}$, we are going to provide examples where knowledge of the Borel field generated by the weak*-topology on $\mathcal{C}$ is insufficient for distinguishing among elements of the partition. Thus, while we assume that S is uncountable, we also give the decision maker a tremendous amount of information (knowledge of the Borel field of $\mathcal{C}$), and this turns out to be insufficient. Our motivation for considering an uncountable state space is that such an assumption allows us to unveil the role of a certain symmetry property, which we believe plays an important role in explaining departures from the the SEU paradigm. Once the structure of the example is understood, it will not be difficult to give examples in the same vein where S is a finite set but the decision maker knows only a finite number of properties (subsets) regarding the set $\mathcal{C}$. We will get back to this in Section 13. We stress, however, that the assumption that S is uncountable is by no means necessary to establish the existence of non-measurable partitions. In fact, as the finite-dimensional examples in Section 9 show, non-measurable partitions of $\mathcal{C}$ easily obtain when S is a finite set.

The first information structure that we consider is not only very natural, but also has an obvious relevance to any theory of decision making. It consists of partitioning the set of measures on Σ so that two measures are in the same cell of the partition if and only if they are associated to the same collection of measure zero events in S. We assume that all the measures in $\mathcal{C}$ are countably addivitive. This is the case, for instance, if a preference relation satisfies, in addition to the axioms 1 to 5 seen above, the axiom of Monotone Continuity (see Appendix D). Formally, for $P, Q \in ca(\Sigma)$ ($ca(\Sigma)$= countably additive measures on Σ), the partition of $ca(\Sigma)$ is defined by the equivalence relation

$$P\mathcal{E}Q \qquad iff \qquad P \ll Q \qquad and \qquad Q \ll P$$

where $\ll$ stands for absolute continuity, and two measures are equivalent if and only if they are mutually absolutely continuous. Informally, the information described by the partition $\mathcal{E}$ corresponds to statements like "The class of measure zero events in Σ is either Φ or Ψ", etc..

Theorem 3 **(see Kechris and Sofronidis** [13]**).** *The partition $\mathcal{E}$ is nonmeasurable.*

As an immediate consequence, we have

Corollary 3. *Let the decision maker's information be given by the quadruple $\{(\mathcal{C}, \mathcal{B}), \mu, \mathcal{E}, \mathcal{B}_{\mathcal{E}}\}$, where $\mathcal{E}$ is the partition produced by the measure equivalence relation. Assume that μ contains a non-atomic part. Then, SEU obtains if and only if 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 (his prior on $\mathcal{C}$ is not concentrated on a single equivalence class), then the decision maker cannot be Bayesian.

In order to see why SEU theory fails, let us try to integrate over $\mathcal{C}$, $\mathcal{C} \subset ca(\Sigma)$, by using a non-atomic prior μ. It is clear that the evaluation of the relative likelihood of (disjoint) events like $Y = \{P \in \mathcal{C} \mid P(A) = 0, A \in \Sigma\}$; and $Z = \{P \in \mathcal{C} \mid P(A) > 0, A \in \Sigma\}$ is necessary to determine the ranking of the acts. This evaluation is equivalent to the evaluation of the indicator function χ_Y. If the decision maker is uncertain about whether or not A has nonzero probability, then suppμ intersects both Y and Z. But, 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: events like Y and Z are *nonmeasurable* with respect to the decision maker's information. Hence, the decision maker cannot evaluate χ_Y. As a consequence, integrating over priors is impossible: 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 μ nor in any other way.

12. Ellsberg's Paradox

12.1. **Ellsberg's three-color urn experiment.** In this section, we consider Ellsberg's three-color urn experiment. Ellsberg's two-urn experiment and Gardenfors and Sahlin's experiment of Section 4 are suitable for similar considerations. 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 black (B) or yellow (Y) in unknown proportions. As it is well-known, the following violation of the SEU paradigm is often observed

$$R \succ B$$

but

$$R \cup Y \prec B \cup Y$$

That is, the decision maker prefers betting on red rather than black, but he prefers betting on "black or yellow" rather than "red or yellow". Notice that the decision maker explicitly receives information about the set of possible configurations of the urn, and that the information is symmetric with respect to the labels B and Y. What is especially interesting in Ellsberg's experiment is that, in correspondence to the symmetry of the information, one typically observes a strong symmetry in the decision maker's table of preferences: one can replace B with Y (and vice versa) at any point in the table of preferences without changing the table itself. We believe that this could hardly be considered an accident.

12.2. **Modeling the symmetry in the information.** Having numbered the balls from 1 to 90, a configuration of the urn is a specification of the color of each ball. Let us denote a configuration by (S, c) to mean that elements of S have been numbered and colored in a certain way. A configuration is possible if and only if the number of red balls equals 30. Now, consider two possible configurations of the urn: one where the number of black balls is 20 and another where the number of black balls is 40. The symmetry with respect to the labels B and Y is evident. To see what this entails in terms of mathematical properties, let us begin by fixing an arbitrary configuration of the urn, say one where the first 30 balls are red, the next 50 are black and the final 10 are yellow. We denote this by $(S, c_0) = \{30, 50, 10\}$. Now, let us consider the configuration $(S, c_1) = \{30, 40, 20\}$. We can think of (S, c_1) as being obtained from (S, c_0) by means of the mapping $g_1 : (S, c_0) \longrightarrow (S, c_1)$ defined by $g_1(s_i) = s_i$, that is g_1 keeps each ball in place, but the balls are now colored according to c_1. Clearly, any configuration of the urn can be described in a similar fashion. We are concerned with the relation existing between the configuration (S, c_1) and the configuration $(S, c_2) = \{30, 20, 40\}$, which is produced by the mapping $g_2 : (S, c_0) \longrightarrow (S, c_2)$. For an arbitrary configuration c_i, let $r : (S, c_i) \longrightarrow (S, r(c_i))$ be the mapping which transforms black balls into yellow balls and vice versa. It is now easy to see that the configurations (S, c_1) and (S, c_2) are linked by the following property: the mappings g_1 and g_2 representing them are such that $g_1 = rg_2r^{-1}$. In such a case, we say that mappings g_1 and g_2 are conjugate. When properly expressed, this is a completly general fact: two configurations of the urn are the relabeling of one another if the mappings representing them are conjugate.

12.3. **A continuous version.** We are going to 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 Y. The set of bets is $\{f \mid f : [0, 1] \longrightarrow \mathbb{R}, f \text{ bounded and } \Lambda\text{-measurable}\}$, where Λ is the usual Borel σ-algebra. The set of possible configurations of the urn is the set of non-atomic measures on $([0, 1], \Lambda)$, which we denote by $\mathcal{N}([0, 1])$. Thus, a configuration $P \in \mathcal{N}([0, 1])$ corresponds to the measure space $([0, 1], \Lambda, P)$, which under our assumptions is a Lebesgue space. By fixing a possibile configuration, say $([0, 1], \Lambda, P_0)$, as a reference point, the Isomorphism Theorem for Lebesgue Spaces (see Appendix B) allows us to identify each configuration $([0, 1], \Lambda, P_i)$ with an invertible measure preserving transformation $g_i : ([0, 1], \Lambda, P_0) \longrightarrow ([0, 1], \Lambda, P_i)$. Thus, the set of all possible configurations of the urn can be identified to the group $G = Aut(P_0)$ of invertible measure preserving transformations of $([0, 1], \Lambda, P_0)$. The notion of symmetry of two configurations of the urn is expressed by the following definition.

Definition 6. *Two configurations, g_1 and g_2 in G, are the relabeling of one another, $g_1 \sim g_2$, if there exists a $r \in G$ such that $g_1 = rg_2r^{-1}$.*

Thus the decision maker's information consists of the partition generated by this equivalence relation along with a prior on G, which specifies that only those measure spaces $([0,1], \Lambda, P_i)$ such that $P_i(R) = 1/3$ should be considered.

Theorem 4 (see Hjorth [12, Theorem 1.2]**).** *The partition associated to the equivalence relation in Definition 6 is nonmeasurable.*

From this, just like in Section 11, it follows that

Corollary 4. *If the decision maker's prior over G contains a non-atomic part and if the prior is not concentrated on a single equivalence class, then SEU fails.*

13. Complements

13.1. **Finitely additive measures.** In Definition 3 of Section 8, we demanded that the decision maker's priors over the set of measures $\mathcal{C}$ be countably additive. As anticipated in the remark following Definition 3, allowing for finitely additive priors only strenghtens our findings. In fact, if a prior μ over $\mathcal{C}$ is finitely additive but not countably additive, then by definition there exists a partition $\mathcal{I}$ and a function $\psi \in B(\mathcal{B})$ such that $\int\limits_{\mathcal{C}} \psi d\mu \neq \int\limits_{\mathcal{C}/\mathcal{I}} \int\limits_{\iota} \psi \mid_\iota d\mu_\iota d\mu'$. It suffices to consider a partition $\mathcal{I} = \{A, \{B_i\}_{i\in\mathbb{N}}\}$ where $\{B_i\}_{i\in\mathbb{N}}$ is a family of disjoint sets for which μ fails countable additivity. Thus, existence of non-measurable partitions is a rather trivial matter when we allow for measures that are only finitely additive.

13.2. **More on the symmetry notion of Section 12.** As we noted, the assumption that the state space be uncountable is by no means necessary to establish the existence of non-measurable partitions. In fact, these appear even when S is just a two-point set. In such a case, the set of measures on S is isomorphic the interval $[0,1]$, and in Section 9 we have seen an explicit example of a non-measurable partition of this set. The introduction in Section 11 of the assumption that S be an uncountable set was motivated by our interest in exploring the notion of symmetric configurations. We saw that that notion of symmetry produces a non-measurable partition. The conceptualization and the result obtained in the case of an uncountable S still turn out to be useful in the case of a finite S. The key observation (see Section 9) is that while we assumed that S was uncountable, at the same time we endowed the decision maker with a tremendous amount of information (the Borel field generated by the weak*-topology). Thus, while keeping exactly the same notion of symmetry, it makes sense to ask what would happen in the case of a finite S if the decision maker had only a limited amount of information. In order to examine this problem, let us suppose that S is a set of n points (endowed with the discrete topology) but that the decision maker's information on $P(S)$, the set of all probabilities on S, consists of only a finite number of properties; for instance, it is described by an algebra generated by a finite number of sets. Having fixed a reference measure, it is easy to see that a measure on S can be represented by means of a matrix on $\mathbb{R}^n$, and that two measures are in the symmetry relation discussed in Section 12 if and only if the corresponding matrices, A and B, are linked by the relation $A = U^*BU$, where U is a (positive) unitary matrix (a matrix U is a unitary matrix if $U^*U = I$, where I denotes the identity matrix). Thus, the ability of distinguishing between non-symmetric measures is the same as that of distinguishing between non-unitarily equivalent matrices. An elementary result in Linear Algebra states that two matrices are unitarily equivalent if and only if they have the same set of eigenvalues. Hence, one needs to able to distinguish between two vectors in $\mathbb{R}^n$. Clearly, this is always possible if the information available contains the Euclidean topology of $\mathbb{R}^n$ (which is countably generated), but this ability might disappear if the information available is coarser. We thus conclude that even in the case of a finite S the notion of symmetry of Section 12 still leads to a failure of SEU theory.

14. Subjectively measurable events

In this section, we focus on decision makers whose information on $(\mathcal{C}, \mathcal{B})$ is given by a partition $\mathcal{I}$ and a prior μ and who, in addition, obey the SEU paradigm conditional on each and every element of the partition. What we are going to say applies without essential modifications to many other types of decision makers (see the remark below). Focusing on those that are conditionally SEU, however, makes the exposition less abstract and is in line with the assumptions that we have made thus far. Given a partition $\mathcal{I}$ of $\mathcal{C}$, decision makers who are conditionally SEU are identified by a collection, $\{\mu_\iota\}_{\iota\in\mathcal{I}}$, of probability measures. In the course of the proof of Theorem 2, we observed that a collection $\{\mu_\iota\}_{\iota\in\mathcal{I}}$

defines an operator $\tilde{T} : \kappa(B(\Sigma)) \longrightarrow \mathbb{R}^{\mathcal{C}/\mathcal{I}}$ by

$$\kappa(f) \longmapsto \tilde{T}(\kappa(f)) \qquad where \qquad \tilde{T}(\kappa(f))(\iota) = \int_{\mathcal{C}} \kappa(f) d\mu_{\iota}$$

When the decision maker faces Ambiguity in the sense of Definition 5, Theorem 2 shows that the collection $\{\mu_{\iota}\}_{\iota \in \mathcal{I}}$ is necessarily non-canonical, and there exist functions in $\kappa(B(\Sigma))$, hence in $B(\Sigma)$, that cannot be evaluated on the basis of the decision maker's information. The complement of this set in $B(\Sigma)$ consists of all those acts (i.e., functions in $B(\Sigma)$) that can be evaluated by the decision maker on the basis of his information. By definition, this set is

$$\Sigma MA = \left\{ f \in B(\Sigma) \mid (a)\ \tilde{T}(\kappa(f))) \in B(\mathcal{B}/\mathcal{I});\ (b) \int_{\mathcal{C}} \kappa(f) d\mu = \int_{\mathcal{C}/\mathcal{I}} \int_{\iota} \kappa(f) \mid_{\iota} d\mu_{\iota} d\mu' \right\}$$

where μ is the decision maker's prior on $\mathcal{C}$. In particular, its subset

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

describes all the events in Σ to which the decision maker can assign probabilities. We will refer to elements of ΣMA as *subjectively measurable acts* and to elements of ΣME as *subjectively measurable events*. The basic properties of the classes ΣMA and ΣME are stated in the next proposition. The proposition also highlights the link between the class ΣME and the class UE of unambiguous events in the sense of [9] and [14].

Proposition 1. *ΣMA is a linear space. Consequently, the class ΣME is non-empty and is a finite λ-system (i.e., is closed under complementation and finite disjoint unions). Furthermore, $UE \subset \Sigma ME$. Finally, there exists a natural measure N on ΣME, defined by*

$$N(E) = \int_{\mathcal{C}} \kappa(\chi_E) d\mu \qquad , \qquad E \in \Sigma ME$$

where μ is the decision maker's prior on $\mathcal{C}$.

Proof. As noted in the proof of Theorem 2, the set

$$\Theta = \left\{ \psi \in B(\mathcal{B}) \mid (a)\ \tilde{T}(\psi) \in B(\mathcal{B}/\mathcal{I});\ (b) \int\limits_{\mathcal{C}} \psi d\mu = \int\limits_{\mathcal{C}/\mathcal{I}} \int\limits_{\iota} \psi \mid_{\iota} d\mu_{\iota} d\mu' \right\}$$

is a linear subspace of $B(\mathcal{B})$. Since a function $f \in B(\Sigma)$ is subjectively measurable if and only if $\kappa(f) \in \Theta$, it follows that the class ΣMA is the set $\kappa^{-1}(\kappa(B(\Sigma)) \cap \Theta)$. From the linearity of κ, it immediately follows that this is a linear subspace of $B(\Sigma)$. Since constant functions always belong to $(\kappa(B(\Sigma)) \cap \Theta)$, the class ΣME always contains $\varnothing$ and S. Moreover, from the fact that ΣMA is a linear space, it follows that ΣME is closed under finite disjoint unions.

If an event $E \in \Sigma$ is unambiguous in the sense of [9] and [14], then $\kappa(\chi_E)$ is a constant mapping on $\mathcal{C}$, that is $\chi_E \in \kappa^{-1}(\kappa(B(\Sigma)) \cap \Theta)$, and $UE \subset \Sigma ME$. Finally, $\chi_E \in \Sigma ME$ implies $\int\limits_{\mathcal{C}} \kappa(\chi_E) d\mu = \int\limits_{\mathcal{C}/\mathcal{I}} \int\limits_{\iota} \kappa(\chi_E) \mid_{\iota} d\mu_{\iota} d\mu'$. Hence, $E \mapsto \int\limits_{\mathcal{C}} \kappa(\chi_E) d\mu$ is an additive set function on ΣME. $\square$

Remark 2. *As anticipated, the notion of subjectively measurable events does not require reference to a system of conditional probabilities. In fact, any functional* V *satisfying condition (7.1), Section 7, produces a mapping* $\kappa(B(\Sigma)) \longrightarrow \mathbb{R}^{\mathcal{C}/\mathcal{I}}$. *We can then define the sets* ΣMA *and* ΣME *just like above. The inclusion* $UE \subset \Sigma ME$ *still holds since constant functions on* $\mathcal{C}/\mathcal{I}$ *are always measurable.*

Of course, if the partition $\mathcal{I}$ is measurable, then SEU obtains, every event in Σ belongs to ΣME, every function in $B(\Sigma)$ is in ΣMA, and the natural set function on Σ is the "average" measure obtained through the integration over priors theorem. If $\mathcal{I}$ is non-measurable and μ is non-atomic, then ΣME is a proper subset of Σ.

14.1. **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. In contrast, events in ΣME are suitable of evaluation, and contracts can be based on those. These considerations provide a link with the problem of unforeseen contingencies as studied in [6]. In fact, the class ΣME is naturally linked to a notion of subjective state space, which is in the spirit of [6]. Given the class ΣME, define a binary relation on $S \times S$ by

$s_1 \approx s_2$ if and only if (modulo sets of N-measure 0)

$$s_1 \in A \Longleftrightarrow s_2 \in A$$

whenever $A \in \Sigma ME$. It is easily seen that $\approx$ is an equivalence relation on S. By definition, the quotient $S/\approx$ consists of all events that the decision maker can evaluate on the basis of his information. The class ΣME and the function $N : \Sigma ME \longrightarrow [0,1]$ of the previous section define a measure-like structure on $S/\approx$. Only functions $S/\approx \longrightarrow \mathbb{R}$ that are measurable with respect to this structure can be evaluated and can, therefore, be included in a contract as possible transfers among the parties. Thus $S/\approx$ can be interpreted as a subjective state space as points in $S/\approx$ represent the only contingencies that the decision maker would include in a contract. Notice that only the functions $S \longrightarrow \mathbb{R}$ which are constant of equivalence classes may correspond to measurable functions on $S/\approx$. Functions $S \longrightarrow \mathbb{R}$ which are not constant of equivalence classes can be thought of as correspondences defined on the quotient as in [8]. By construction, these correspondences do not admit measurable selections.

15. De Finetti re-considered

The results obtained in this paper make it clear that de Finetti's argument relies on the implicit assumption that the information available to the decision maker is good information. Formally, de Finetti's argument implies that the decision maker has enough information to integrate all the functions on $B(\Sigma)$. It follows from Section 14 that this is the case if and only if $B(\Sigma) = \kappa^{-1}(\kappa(B(\Sigma)) \cap \Theta)$ which occurs if and only if $\kappa(B(\Sigma)) = \Theta$. In turn, this occurs if and only if the decision maker's information corresponds to Risk and not to Ambiguity (Definition 5).

Appendix A: Rationality $\Longrightarrow$ SEU

The proposition contained in this appendix fills in the details of the argument outlined in Section 2 that Rationality $\Longrightarrow$ SEU. The proposition is well-known. Its inclusion is motivated only by the need of making this paper reasonably self-contained.

Proposition 2. *Let $I : B(\Sigma) \longrightarrow \mathbb{R}$ be additive, monotone and such that $I(\mathbf{1}) = 1$. Then, there exists a unique finitely additive probability measure $P \in ba(\Sigma)$ such that for every $f \in B(\Sigma)$*

$$I(f) = \int_S f dP$$

Proof. Step 1: For any $f \in B(\Sigma)$, $I(-f) = -I(f)$.

By additivity, $1 = I(\mathbf{1}) = I(\mathbf{1} + \mathbf{0}) = I(\mathbf{1}) + I(\mathbf{0})$, which implies $I(\mathbf{0}) = 0$. From $0 = I(\mathbf{0}) = I(f - f) = I(f) + I(-f)$, we conclude that $I(-f) = -I(f)$ for any $f \in B(\Sigma)$.

Step 2: For any rational number $\lambda \in \mathbb{Q}$ and any $f \in B(\Sigma)$, we have $I(\lambda f) = \lambda I(f)$.

For any $f \in B(\Sigma)$ and any natural number $n \in \mathbb{N}$, additivity implies $I(nf) = nI(f)$. By this and Step 1, we then conclude that $I(zf) = zI(f)$ for any relative integer $z \in \mathbb{Z}$. Moreover for every $z \in \mathbb{Z}\backslash\{0\}$,

$$I(f) = I(\frac{z}{z}f) = zI(\frac{1}{z}f) \qquad \Longrightarrow \qquad \frac{1}{z}I(f) = I(\frac{1}{z}f)$$

If $\lambda \in \mathbb{Q}$, then $\lambda = p/q$ with $p, q \in \mathbb{Z}$. By combining the previous observations, we see that

$$I(\lambda f) = I(\frac{p}{q}f) = pI(\frac{1}{q}f) = \frac{p}{q}I(f) = \lambda I(f)$$

Step 3: I is sup-norm continuous

For $f, g \in B(\Sigma)$, let $\{r_n\} \subset \mathbb{Q}$ be a sequence of rationals which converges from above to $\|f - g\|_\infty$. From $f = g + (f - g)$ and $g = f + (g - f)$, by using monotonicity, additivity, $1 = I(\mathbf{1})$ and the result in Step 2, we see that for any n

$$\begin{aligned} I(f) &\leq I(g) + I(\|f - g\|_\infty) \leq I(g) + r_n \\ I(g) &\leq I(f) + I(\|f - g\|_\infty) \leq I(f) + r_n \end{aligned}$$

Hence, for any n

$$|I(f) - I(g)| \leq r_n$$

By taking the limit for $n \to \infty$, we have

$$|I(f) - I(g)| \leq \|f - g\|_\infty$$

that is the sup-norm continuity of I.

We can now conclude the proof. By combining Step 2 and Step 3, we see that I is homogeneous, that is $I(\lambda f) = \lambda I(f)$ for all $f \in B(\Sigma)$ and for all $\lambda \in \mathbb{R}$. Along with addivity, this implies that I is linear. By Step 3, I is also sup-norm continuous. By the Riesz representation theorem, there exists a unique $P \in ba(\Sigma)$ such that for every $f \in B(\Sigma)$

$$I(f) = \int_S f dP$$

Finally, monotonicity of I implies that P is non-negative, and $I(\mathbf{1}) = 1$ implies $P(S) = 1$. □

Appendix B: 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.

Given two measurable spaces, $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$, a mapping $X_1 \longrightarrow X_2$ is called a Borel isomorphism if it is a bijection and is bimeasurable.

Borel isomorphism theorem (see [19, Theorem 3.3.13]**):** Any two uncountable standard Borel spaces are Borel isomorphic.

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

Given two measure spaces, $(X_1, \mathcal{B}_1, m_1)$ and $(X_2, \mathcal{B}_2, m_2)$, a measurable mapping $T : X_1 \longrightarrow X_2$ is a *measure preserving transformation* if for all $E \in \mathcal{B}_2$ we have

$$m_1(T^{-1}(E)) = m_2(E)$$

If T is bijective and its inverse T^{-1} is also measure-preserving, then T is an *invertible measure-preserving transformation.* Two measure spaces, $(X_1, \mathcal{B}_1, m_1)$ and $(X_2, \mathcal{B}_2, m_2)$, are isomorphic if there exists an invertible measure preserving transformation $T : X_1 \longrightarrow X_2$.

Isomorphism of Lebesgue Spaces (see [20, Theorem 2.1]**):** Any two Standard Lebesgue spaces are isomorphic.

Appendix C: Conditional measures

Let $(\mathcal{C}, \mathcal{B}, \mu)$ be a measure space, let $\mathcal{I}$ be a partition of $\mathcal{C}$ (modulo μ-measure 0 events) and let $\mathcal{C}/\mathcal{I}$ denote the quotient space endowed with the finest σ-field that makes the canonical projection measurable. We recall the following definition.

Definition 7. *A canonical system of conditional measures associated to the partition $\mathcal{I}$ is a family of measures $\{\mu_\iota, \iota \in \mathcal{I}\}$, with the following properties*

(i) for any $A \in \mathcal{B}$, the set $A \cap \iota$ is measurable in ι for almost all $\iota \in \mathcal{C}/\mathcal{I}$ and the function $\mu_\iota(A \cap \cdot) : \mathcal{C}/\mathcal{I} \to \mathbb{R}$ is measurable; and

(ii) for any $A \in \mathcal{B}$,

$$\mu(A) = \int_{\mathcal{C}/\mathcal{I}} \mu_\iota(A \cap \iota) d\mu'$$

where μ' is the image measure (pushforward) of μ under the canonical projection $\pi : \mathcal{C} \longrightarrow \mathcal{C}/\mathcal{I}$.

Appendix D: A Polish setting

In combination, the two assumptions below guarantee that the set of measures $\mathcal{C}$ in Theorem 1 is a Polish space.

Standard State Space: The measurable space (S, Σ) is a standard Borel space.

Let $\succsim$ be a preference relation satisfying axioms 1 to 5. Let $\succsim^*$ denote the unambiguous preference relation ([9], Sec. B.3) associated to $\succsim$.

Axiom of Monotone Continuity (see [9]**):** 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 $y \succ^* x A_{\bar{n}} z$.

The Axiom of Monotone Continuity is equivalent to the property that all the measures in Theorem 1 are countably additive ([9], Sec. B.3).

Theorem 5. *Let (S, Σ) be a standard Borel space. A preference relation $\succsim$ on $\mathcal{A}$ satisfies axioms 1 to 5 and the axiom of Monotone Continuity iff the functional I representing it factors as $I = V \circ \kappa$, where*

(i) $\kappa : B(\Sigma) \longrightarrow B(\mathcal{B})$ is defined by $\kappa : f \longmapsto \psi_f$, where $\psi_f : \mathcal{C} \longrightarrow \mathbb{R}$ is given by $\psi_f(P) = \int f dP$, $P \in \mathcal{C}$.

(ii) κ is linear and sup-norm to sup-norm continuous;

(iii) $\mathcal{C}$ is a Polish space;

(iv) $V : B(\mathcal{B}) \longrightarrow \mathbb{R}$ is monotone and comonotonic additive (hence, sup-norm continuous).

In the course of the proof, we will denote by $\sigma(ba(\Sigma), Y)$ the weak topology on $ba(\Sigma)$ induced by a set of mappings Y.

Proof. From Theorem 1, we know that $\mathcal{C}$ is a weak*-compact subset of $(ba, \sigma(ba(\Sigma), B(\Sigma)))$. By the Axiom of Monotone Continuity, all the probabilities in $\mathcal{C}$ are countably additive. By the assumption that (S, Σ) is standard Borel, it follows that all the probabilities in $\mathcal{C}$ are regular. If we replace the topology $\sigma(ba(\Sigma), B(\Sigma))$ with $\sigma(ba(\Sigma), C_b(S))$, then $\mathcal{C}$ remains compact because the new topology is weaker than the original one. In particular, $\mathcal{C}$ is closed. Finally, (S, Σ) standard implies that the space $\mathcal{P}(\Sigma)$ (=regular Borel measures on Σ) is Polish in the topology $\sigma(\mathcal{P}(\Sigma), C_b(S))$, and we conclude that $\mathcal{C}$ is Polish as well.

Next, define κ like in the statement of the theorem. Thus, $\kappa : B(\Sigma) \longrightarrow \mathbb{R}^{\mathcal{C}}$ and κ is clearly linear. Since κ is defined (pointwise) in exactly the same way as in Theorem 1, the factorization established there obtains here as well. Since all the measures in $\mathcal{C}$ are bounded and countably additive, the Monotone Convergence Theorem implies that κ is normal, that is

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

Let $E \in \Sigma$, and let χ_E denote the indicator function of the set E. 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 $\sigma(\mathcal{P}(\Sigma), C_b(S))$ ([1], Lemma 14.16); that is, $\kappa(\chi_E) \in B(\mathcal{B})$ for all $E \in \Sigma$. If $h \in B(\Sigma)$ is a simple function, then h can be written as a finite linear combination of indicator functions, and linearity of κ along with the previous conclusion imply that $\kappa(h) \in B(\mathcal{B})$. Finally, if $f \in B(\Sigma)$ is any function, then there exists a sequence of simple functions $\{f_n\} \subset B(\Sigma)$ such that $f_n \nearrow f$, and normality of κ implies that $\kappa(f) \in B(\mathcal{B})$. We conclude that when $\mathcal{C}$

is equipped with the Polish topology $\sigma(\mathcal{P}(\Sigma), C_b(S))$, the linear mapping κ takes $B(\Sigma)$ into a subset of $B(\mathcal{B})$. Moreover, κ is sup-norm to sup-norm continuous as a consequence of the inequality

$$\begin{aligned} \|\kappa(f) - \kappa(g)\|_\infty &= \sup_{P\in\mathcal{C}} \left| \int f dP - \int g dP \right| \le \sup_{P\in\mathcal{C}} \int |f-g| \, dP \\ &\le \sup_{P\in\mathcal{C}} \int \sup_{s\in S} |f-g| \, dP = \|f-g\|_\infty \end{aligned}$$

Finally, V is defined exactly as in Theorem 1. $\square$

References

[1] Aliprantis C.D. and K.C. Border (2006), Infinite Dimensional Analysis, 3rd ed. Springer-Verlag.

[2] Amarante M (2009), Foundations of Neo-Bayesian Statistics, *Journal of Economic Theory* **144**, 2146-73.

[3] Anscombe F.J. and R.J. Aumann (1963), A definition of subjective probability, *Annals of Mathematical Statistics* **34**, 199-205.

[4] Billingsley P. (1995), Probability and Measure, Wiley.

[5] Cornfeld I.P., S.V. Fomin and Y.G. Sinai (1982), Ergodic Theory, Springer-Verlag.

[6] Dekel E., B. Lipman and A. Rustichini (2001), Representing Preferences with a Unique Subjective State Space, *Econometrica* **69**, 891-934.

[7] Gardenfors P. and N-E Sahlin (1982), Unreliable Probabilities, Risk-taking, and Decision Making, *Synthese* **53**, 361-86.

[8] Ghirardato P. (2001), Copying with Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty, *Economic Theory* **17**, 247-76.

[9] Ghirardato P., F. Maccheroni and M. Marinacci (2004), Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory* **118**, 133-173.

[10] Ghirardato P., F. Maccheroni, M. Marinacci and M. Siniscalchi (2003), Subjective Foundations for Objective Randomization: A New Spin on Roulette Wheels, *Econometrica* **71**, 1897-1908.

[11] Gilboa I. and D. Schmeidler (1989), Maxmin expected utility with a non-unique prior, *Journal of Mathematical Economics* **18**, 141-53.

[12] Hjorth G. (2000), Classification and orbit equivalence relations, American Mathematical Society.

[13] 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.

[14] Nehring K. (2001), Ambiguity in the context of probabilistic beliefs, UC Davis, mimeo.

[15] Phelps R. R. (1966), Lectures on Choquet theorem, van Nostrand.

[16] Rokhlin, V.A. (1949), On the Fundamental Ideas of Measure Theory, Translations of the AMS, 71.

[17] Savage L.J. (1972), The Foundations of Statistics, Dover.

[18] Schmeidler D. (1989), Subjective probability and expected utility without additivity, *Econometrica* **57**, 571-87.

[19] Srivastava, S.M. (1998), A Course on Borel Sets, Springer-Verlag GTM.

[20] Walters P., F. W. Gehring and K. A. Ribet (2000), An Introduction to Ergodic Theory, Springer.

Département de sciences économiques, Université de Montréal et CIREQ
E-mail address: massimiliano.amarante@umontreal.ca