Columbia University

Department of Economics
Discussion Paper Series

States, Models and Unitary Equivalence I: Representation Theorems and Analogical Reasoning

Massimiliano Amarante

Discussion Paper No.: 0405-10

Department of Economics
Columbia University
New York, NY 10027

December 2004

States, Models and Unitary Equivalence I: Representation Theorems and Analogical Reasoning

Massimiliano Amarante*
Columbia University, NY†

October 28, 2004

Abstract

I show that virtually any model of decision making under uncertainty is associated to a certain structure. This contains three fundamental ingredients: (1) The domain of the acts; (2) Another set, which is called the set of models for the decision maker; and (3) The decision maker's information about the set of models (an algebra of subsets of the set of models). A consequence of this finding is that that the decision maker's choices can be viewed as the outcome of a two-stage process. First, the set of acts is mapped into a system of hypothetical bets on the set of models. Then, the latter are ranked by the decision maker. I show that this procedure can be thought of as describing a general form of analogical reasoning. I also observe that the appearance of two different sets implies that the decision maker is uncertain about two different objects and that he may receive information about any of them. In particular, information about the set of models affects the decision maker's ranking of the available alternatives. In the sequel to this paper, I show that certain natural information structures lead to an inherent inability of assigning probabilities on the domain of the acts. In a formal sense, their properties describe the idea of Knightian Uncertainty.

*I benefited from comments and suggestions from Larry Epstein, Paolo Ghirardato, Tzachi Gilboa, Fabio Maccheroni, Massimo Marinacci, Rich McLean, Ket Richter, Aldo Rustichini, Marco Scarsini, Paolo Siconolfi, Max Stinchcombe, William Thomson, Stan Wellisz and Bill Zame. I also wish to thank seminar participants at University of Iowa, RUD 2003, Cal-Tech, University of Minnesota, University of Texas-Austin, Columbia University, University of Rochester, Rutgers University, University of Turin, Ohio State University, Columbia University Workshop on Multiple Priors.

†Department of Economics, Columbia University, 1019 IAB, 420 W 118th street, NY, NY 10027. Email: ma734@columbia.edu.

1 Introduction to States, Models and Unitary Equivalence

This is the first of a series of papers under the heading *States, Models and Unitary Equivalence*. The goal of this work is to lay the foundations of a theory of decision making under uncertainty, where Information plays an explicit and substantial role. Ultimately, the theory is going to produce (normative) statements of the sort, "If a decision maker has a certain information x, then he should display behavior of type $b(x)$", or (positive) statements of the sort, "if a decision maker displays behavior of type y, then we can think of him *as if* he followed a certain set of rules and his information were of type $I(y)$". We shall see that, in many instances, the content of such statements will be rather novel. Among other things, we are going to see that (1) the old, but still vague, idea of Knightian Uncertainty can be formally described by means of the properties of certain information structures; and that (2) behavior not conforming to the classical Subjective Expected Utility (SEU) paradigm may emerge as a "rational" response in the presence of those information structures.

The theory of decision making under uncertainty was first systematized by the work of Savage [31]. Later, Prospect Theory [30] and the work of Schmeidler [32] and Gilboa and Schmeidler [16] have brought about new ways of looking at the problem. In particular, the widely observed violation of the Independence Axiom (Ellsberg [7]) has been accommodated in the context of these theories. However, in my opinion, a fully satisfactory explanation is still lacking. By itself, this is a good enough reason for a thorough reconsideration of the foundations of the theory. Moreover, the recent works on Unforeseen Contingencies (Dekel, Lipman and Rustichini [4]), on Dynamic Choice (Epstein and Schneider [8], Gul and Pesendorfer [19]), on Case-Based decision theory (Gilboa and Schmeidler [17], [18]) and on Ambiguity (Epstein and Zhang [9], Ghirardato, Maccheroni and Marinacci [12], Nehring [29]) have brought about several new ideas which make such a reconsideration all the more pressing. I believe that the explicit recognition of the role played by Information in decision problems is a necessary prerequisite for a successful systematization of these ideas. I hope to accomplish this with the present work.

My exposition is motivated by three themes, which recur often in recent works on decision theory. These are: (1) Analogical Reasoning; (2) Knightian Uncertainty; and (3) Interpretation of Multiple Prior models. The common thread running through these three themes is the idea of Information. Here, I limit myself to summary hints. Needless to say, each of these themes will be analyzed in great detail as the work progresses.

There is little doubt that analogical reasoning is one of the most prominent forms of human thinking. Yet, classical theories of decision making neither recognize its role explicitly nor provide a formalization of the notion. One of my goals is to reach a formal definition and to study its implications for a theory of decision making under uncertainty. Important work in this direction has recently been done by Gilboa and Schmeidler ([17], [18]; see Section 4).

Informally, the idea of Knightian Uncertainty designates those situations characterized by the decision maker's inherent inability to assign probabilities to the various events. While thinking about these kind of situations has proved remarkably useful, no formal characterization has been provided. I will do so in the present work. I will show that certain information patterns lead, in a formal sense, to the inability of using probabilities on the domain of the acts.

Many decision theorists favor Multiple Prior models. In Marinacci's words, " ... 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 his limited information" (Marinacci [27]). While the statement is very suggestive, it has no formal counterpart within the existing theories. As a matter of fact, it is the concept of "information" that does not find an appropriate formalization in those theories. In this work, I explicitly introduce the concept of information structure, and study how the decision maker's choices vary as a function of the underlying information structure. By doing so, I will be able, among other things, to justify statements like Marinacci's on a formal ground.

Roughly, each of these points will be dealt with in a separate paper in the order given here.

To conclude this introduction, a few words about the series' heading are in order. The building block of my analysis is the recognition that virtually any problem of decision making under uncertainty is associated to a certain structure (Sections 6 and 7 in this paper). This structure contains three fundamental ingredients. The first is the domain of the acts. After Savage, this is called the set of states. The second is another set that can be thought of as the collection of all the risky descriptions associated with the set of states (Sections 6 to 8). This is called the set of models for the decision maker. The motivation for such a terminology is provided by the construction in Section 5. It is a consequence of this finding that the decision maker's choices can be viewed as the outcome of a process involving both the set of states and the set of models, and which depends on the decision maker's understanding of the latter. This brings us to our third ingredient, which is a description of the decision maker's information about the set of models. Formally, it corresponds to the sub-algebra of subsets generated by his information.

In the sequel to this paper, I study the decision maker's behavior as a function of the information structure assigned on the set of models. In doing so, I show that certain natural information structures lead to an inherent inability of assigning probabilities on the domain of the acts. Hence, they are naturally associated to the idea of Knightian Uncertainty. All of these information structures display a key property, namely, they produce a partition on the set of models which is ergodic. The archetype of such a phenomenon is provided by the unitary equivalence of normal operators on a Hilbert space, which is also the main tool in some of the proofs.

2 What this paper is about

2.1 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 $I : F^+(S) \to \mathbb{R}_+$ having a certain form”.[1]

$$\begin{array}{ccc} F^+(S) & & \\ & I \quad \searrow & \\ & & \mathbb{R}_+ \end{array}$$

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 this paper, I am going to show (Sections 6 and 7) that, in essentially any axiomatic model, the functional I can be thought of as consisting of two parts as in the diagram below.

$$\begin{array}{ccc} F^+(S) & \stackrel{\kappa}{\longrightarrow} & F^+(M) \\ & I \quad \searrow & \downarrow V \\ & & \mathbb{R}_+ \end{array}$$

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. 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 the role played by “Information” in decision problems.

[1] Here $F^+(S)$ denotes the set of nonnegative (measurable) real-valued functions on S, and R_+ denotes the nonnegative real numbers.

2.2 Analogical reasoning

On the front of novel interpretations, the results contained in this paper allow us to link existing axiomatic theories to the idea of analogical reasoning. In Section 5.1, I give a formal definition of what it means to solve a problem by analogy with another. The idea is simple and intuitive. Suppose that a decision maker has already ranked a set of alternatives B. Now, he faces the problem of ranking a new set of alternatives A. The idea is that he would map the new alternatives into the old ones, and then use the past ranking to order the new ones.

Of course, the process of analogical reasoning can be much more sophisticated than just described. In general, we would expect a decision maker to use all of his past experience to solve the problem at hand. In Section 5.3, I build a model of decision making that tries to capture this intuition, and show that it is described by the diagram in Figure 2. That is, it displays the same structure as any axiomatic model of decision making. Thus, those models can now be reinterpreted in terms of analogical reasoning, and new insights can be gained from this perspective.

2.3 Uncertainty and information

Figure 2 states that the process of decision making involves two sets: the domain of the acts S and another set M. What is this set? A remarkable feature of the representation theorems of Sections 6 and 7 is the result that M can always be represented in the same way. Namely, it is a set of probabilistic descriptions of the domain of the acts (see also Section 8). 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.

In the sequel to this paper, I argue that the simultaneous presence of S and M reflects the existence of two conceptually different types of uncertainty confronting the decision maker: uncertainty about which state obtains and uncertainty about "how the world works". In doing so, I depart from the classical point of view (Savage [31], Ch. 2) which focuses only on uncertainty of the first type. 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. Hence, one is led to inquire about how information about M affects the decision maker's ranking of the alternatives. This study is very much in the spirit of the interpretation in terms of analogical reasoning proposed above. In a way, it corresponds to the study of how a decision maker's experience/knowledge affects his perception of the problem he faces.

2.3.1 States, Models and Unitary Equivalence II

Ellsberg's experiment (Section 10 in this paper) is a stripped down example that brings to light the fact that information about the set M might affect the decision maker's choices. The general problem is the main theme in part II of this work. The main finding is that certain natural information structures on M produce an inherent inability of using probabilities on S. In a precise sense, these capture the vague idea of Knightian Uncertainty, which thus becomes suitable of formal analysis. In this context, behavior of Ellsberg's type is shown to emerge as the outcome of purely informational phenomenona.

3 The setting

Before moving to the main issues dealt with in this work, a few words about the setting of decision making under uncertainty are in order. The goal is not to refresh the readers' memory, but rather to make some observations about the nature of the objects of study. Specifically, I am going to argue that, if we aim for theories of decision making that are truly testable, then there can be no ambiguity as to what those objects are. We shall see that important consequences stem from the considerations of this section.

The setting of decision making under uncertainty is as follows. An individual, the decision maker, has to rank alternatives from a set $\mathcal{F}$. Given an element s in a set S, each alternative $f \in \mathcal{F}$ produces an outcome, $f(s) = x \in X$. Hence, elements in $\mathcal{F}$ can be viewed as mappings $S \to X$, and the decision maker has to rank those. Given the decision maker's preference relation over outcomes, the problem is nontrivial if the decision maker does not know which element in S obtains. Summarizing, we have the following objects:

(A) A set S. In Savage's terminology, this is the set of states.

(B) A set of outcomes X.

(C) A set of alternatives $\mathcal{F}$. This is a collection of mappings $S \to X$, and elements in $\mathcal{F}$ are called acts.

In all axiomatic theories, such a description is enriched by introducing a fourth object. Namely,

(D) A fixed field, Σ, of subsets of S.

Most theories of decision making are "choice-based". That is, the theory's only *datum* is the decision maker's ranking of the alternatives and, in turn, all the theory's prescriptions translate into assessments about the decision maker's ranking. Since such a ranking is observable, this guarantees that the theory is testable. If we adhere to this principle, then there can be no ambiguity as to what the above objects are.

To begin, the set of states S should be deprived of any metaphysical meaning. S should be viewed as the domain of the acts. Both the decision maker and any outside observer have to be able to agree as to what S is. A similar consideration can be made about X and $\mathcal{F}$.[2]

Of course, this is not to say that the decision maker's ranking of the alternatives depends on S alone. For instance, if we think in terms of reasoning by

[2] The necessity that S be objectively given has been stressed in a recent paper by Karni [24].

analogy, we must admit that the decision maker might have information regarding another set, say C, and know that C and S are related in a certain way. Then, he can use that information when it comes to ranking the elements in $\mathcal{F}$. It goes without saying that phenomena of this sort are not only possible, but also that a comprehensive theory of decision making should certainly account for them. I would like to stress, however, that both the existence of a certain relation between C and S (known to the decision maker) and the decision maker's information about C are, generally speaking, non-observable.

The final observation regards the field Σ. In axiomatic theories, the role of Σ is to transform acts into measurable functions (by using a utility function on the outcome space). This is clearly a necessary step for theories aiming for some kind of an expected utility criterion. Here, the issue is about the meaning of Σ. One view might be that Σ reflects the decision maker's information about S. This is patently illegitimate if we want the theory to be testable as such information is unobservable. If anything, it should be derived from the decision maker's ranking.[3] However, Σ can be given on objective meaning: it can be defined as the coarsest field which makes all the acts measurable. This is completely determined by $\mathcal{F}$, which is objectively given. Moreover, this does not imply anything about the decision maker's information about S. Throughout this work, we will stick with this interpretation.

These considerations about the meaning of Σ might appear redundant. To my knowledge, no axiomatic theory has ever treated, at least explicitly, Σ as reflecting the decision maker's information about S. Nevertheless, some confusion seems to be present in various debates. For instance, consider the debate surrounding the three-color urn experiment of Ellsberg. A well-known argument dismisses the choices not conforming to the SEU paradigm as irrational since they seemingly violate an elementary form of consistency of logical reasoning. By adhering to this argument, however, one implicitly assumes that the decision maker uses a separated field on the domain of the bets. Specifically, one presumes that the set of blue balls and the set of green balls are measurable events from the viewpoint of the decision maker (and, as such, suitable for evaluation by means of a probability measure), while there is nothing guaranteeing that this is indeed the case. A similar observation can be made for the no-arbitrage argument of de Finetti or even for the Bayesian argument stating that a decision maker described by a prior over a set of priors is equivalent, from an observational viewpoint, to a decision maker described by a single prior. Hoping that the examples have at least convinced the reader that I might have a point, I will leave it at that, and return to it in subsequent sections.

In the next section, I am going to set $X = \mathbb{R}$, the real line. This is fully justified when one imposes axioms on preferences which guarantee the existence of a utility function on the outcome space. I will do so in Section 6. In addition, for reasons that will become clear later in the paper (Section 9), I am going to define Σ as a σ-algebra, and not simply an algebra.

[3] The work on Ambiguity (Epstein and Zhang [9], Ghirardato, Maccheroni and Marinacci [12], Nehring [29], Stinchcombe [35]) can be interpreted as replying exactly to this need.

4 Reasoning by Analogy

Analogy is the recognition that one thing A (a phenomenon, a problem, etc.) is like another thing B and that, therefore, consequences (inferences, explanations, solutions, etc.) that can be drawn from A can be drawn from B as well. Analogy is a reasoning process that is so pervasive in human life that can be viewed as one of the cornerstones of human thought. In the Dictionary of Philosophy of Mind [6], Analogy is recognized as, "... an important kind of thinking contributing to such cognitive tasks as explanation, planning and decision making (Paul Thagard)". Because of this, it is not surprising that the concept of Analogy has been an object of philosophical reflection since ancient times.[4]

As of today, research on the concept of Analogy and on the process of analogical reasoning has become increasingly important in the study of Artificial Intelligence (see, for instance, [11]). Again, this is not surprising given the nature of the subject. What is surprising is that classical theories of decision making under uncertainty do not explicitly recognize the role played by analogical reasoning.

Theories of decision making under uncertainty are concerned with explaining/guiding the behavior of individuals who have to choose a course of action in an uncertain environment. It is indisputable that, in a large variety of situations, individuals perceive this as a "new" problem, in the sense that they do not have a "ready-made" solution for it. In such situations, individuals tend to rely on their past experience, that is, they recall solutions that they gave in the past and try to "adapt" them to the new circumstances. It is precisely these types of processes that are unaccounted for by classical theories.

In recent years, Gilboa and Schmeidler started a research program (see, for instance, the comprehensive [17] and the more recent [18]) aiming to fill, at least in part, this gap. Their Case-Based Theory of Decisions can be viewed as a specialization to a decision theoretic setting of the idea of Case-Based Reasoning. This is founded on the intuition that the problems one faces are often similar to problems encountered in the past and, therefore, that past solutions may be of use in current situations. The idea of Case-Based Reasoning becomes especially powerful when the reasoning method is coupled with a learning paradigm: solutions given to new problems update the case base thus producing new strategies to attack new problems. Strictly speaking, Case-Based Reasoning refers to problems based on single-domain cases.[5] As such, it is a special case of Analogical Reasoning that allows for the use of past cases from different domains and goes even beyond that.

In the next section, I propose a definition of Reasoning by Analogy that is tailored to a setting of decision making under uncertainty. Building on this definition, I construct a model of decision making which accounts for the possibility that a decision maker might choose a course of action based on past

[4] To give just one example, reflections on the concept recur several times in St. Thomas from the Summa Theologica to the De potentia to the De veritates.

[5] Roughly speaking, a set of cases pertain to a single domain if they are describable by means of the same set of parameters.

or even on only potential cases from different domains. The model contains several ingredients that combine to give rise to a certain structure. In Section 6, I show that one of the most general axiomatic models of decision making we know of (see [12]) displays the same structure as the model built on the idea of analogical reasoning. In a way, that structure had always been there, and the task accomplished by Theorem 2 (Section 6) is that of bringing that structure to light. Later, I show that the result holds for axiomatizations weaker than that in [12].

5 Models and Analogies

5.1 Simple analogies

Suppose that a decision maker has to rank a set of bets, $\mathcal{F}$, regarding the meteorological conditions in Siberia (S) over the month of July in a certain year. Lacking any direct experience of the weather in Siberia, our decision maker might try to "translate" this problem into a problem he is more familiar with. For instance, the meteorological conditions over the same month in his own state, which, we suppose, is California (C). In other words, let us assume that the decision maker has some information about C, described by a certain σ-algebra $\mathcal{C}$, and that, on the basis of this information, he is able to rank all the bets in

$$F^+(C) = \{f : C \to \mathbb{R}_+ \mid f \text{ bounded and } \mathcal{C}\text{-measurable}\}$$

Let us denote by $\succsim_C$ the decision maker's ranking of the bets in $F^+(C)$. Now, our decision maker wants to use this knowledge, and possibly the knowledge that Siberia and California are at a different latitude, longitude, that they are both on the Pacific, etc., to solve the problem he faces, namely that of ranking the bets in $\mathcal{F}$.

To begin, like in the previous section, define Σ as the coarsest σ-algebra which makes all the bets in $\mathcal{F}$ measurable, and (to simplify the exposition) set

$$\mathcal{F} = F^+(S) = \{f : S \to \mathbb{R}_+ \mid f \text{ bounded and } \Sigma\text{-measurable}\}$$

Next, we want to formalize the idea that the problem of ordering $F^+(S)$ can be "translated" into the problem of ordering $F^+(C)$ (which has already been solved by the decision maker with the ordering $\succsim_C$). The task boils down to identifying two requirements. First, it must be the case that each bet $f_i \in F^+(S)$ can be identified to a bet $\phi_i \in F^+(C)$. In other words, there must exist a mapping $\nu : F^+(S) \to F^+(C)$. Second, it must be the case that, modulo the renaming of the bets produced by the mapping ν, the objects $F^+(S)$ and $\nu(F^+(S)) \subset F^+(C)$ are the same. This means that if there exists a certain relation between any two bets $f, g \in F^+(S)$, then this relation must be preserved once the mapping ν is applied. That is, it must be the case that the mapping ν preserves whatever structure is associated with $F^+(S)$. In our context, $F^+(S)$ is an affine space of measurable mappings. Hence, the requirement that ν be structure-preserving translates into the demand that ν displays the following properties

(i) affinity: $\nu(\alpha f + \beta g) = \alpha\nu(f) + \beta\nu(g), \qquad \alpha, \beta \in \mathbb{R}_+$
(ii) normality: $f_n \nearrow f \Longrightarrow \nu(f_n) \nearrow \nu(f), \qquad n \in \mathbb{N}.$

That is, if a bet $h \in F^+(S)$ is a combination of two other bets, $f, g \in F^+(S)$, then we want $\nu(h)$ to be an analogous combination of the bets $\nu(f)$ and $\nu(g)$; and if a collection of bets $\{f_n\}$ in $F^+(S)$ approximates a bet $f \in F^+(S)$, then we want the collection $\{\nu(f_n)\}$ to approximate the bet $\nu(f)$. Mappings like ν are called kernels or generalized transition probabilities, and will play an important role throughout this work (for more about kernels, see Appendix A.5.1).

Once $F^+(S)$ and $F^+(C)$ are made to correspond to each other by means of a kernel $\nu : F^+(S) \to F^+(C)$, then the problem of ordering $F^+(S)$ can be solved by setting

$$f \succsim g \qquad iff \qquad \nu(f) \succsim_C \nu(g)$$

We summarize the content of this discussion in the following definition.[6]

Definition 1 *Let (S, Σ) and $(C, \mathcal{C})$ be two measurable spaces. Let d be a decision maker who displays choices $\succcurlyeq_S$ and $\succcurlyeq_C$ on $F^+(S)$ and $F^+(C)$, respectively. We say that the problem $(F^+(C), \succsim_C)$ is a model for ranking the bets in $F^+(S)$ if there exists a kernel $\nu : F^+(S) \to F^+(C)$ such that*

$$f \succcurlyeq_S g \qquad iff \qquad \nu(f) \succcurlyeq_C \nu(g)$$

We say that the problems $(F^+(C), \succsim_C)$ and $(F^+(S), \succsim_S)$ are analogous if the kernel is a bijection.

The definition makes it clear that the concepts of *model* and *analogy* pertain to a decision maker. In other words, one can exhibit two decision makers, d_1 and d_2, for whom the problems on S and C are analogous but the analogy is realized by different kernels as well as a third decision maker for whom there exist no analogy between the two problems.

5.2 Kernels and models. An example

As a general matter, the possibility of solving a problem, say ordering $F^+(S)$, by analogy with another problem, say $(F^+(C), \succsim_C)$, corresponds to the existence of an affine mapping $\nu : F^+(S) \to \mathbb{R}_+^C$ with the property that $range(\nu) \subseteq F^+(C)$. As the reader has probably already realized, the existence of such a mapping depends crucially on the relation between the two σ-algebras $\mathcal{C}$ and Σ or, equivalently, between $\mathcal{C}$ and the set of bets $\mathcal{F}$ offered to the decision maker.

It is important that the reader keep in mind that the issue, far from being merely technical, is a very substantial one. The σ-algebra $\mathcal{C}$ on C describes (by assumption) the decision maker's understanding of C. This is the same as saying that the decision maker's knowledge is described by $F^+(C)$. Now, suppose that the decision maker attempts to translate his problem into $(F^+(C), \succsim_C)$, and that this produces a certain (non-constant) affine mapping $\nu : F^+(S) \to \mathbb{R}_+^C$.

[6] Our definition is in the spirit of [10]. The latter has been criticized in [3]. See also [28] and [11] for more on the debate.

Suppose further that $\mathcal{C}$ is so coarse that this mapping does not satisfy the condition $range(\tilde{\nu}) \subset F^{+}(C)$. In a manner of speaking, what happens is that $F^{+}(C)$ is not big enough to accommodate all the functions in $F^{+}(S)$. This situation has a very transparent meaning. Knowledge of $(F^{+}(C), \succsim_C)$ is not enough to solve the problem at hand. Equivalently, $(F^{+}(C), \succsim_C)$ cannot be a model for ordering the bets in $F^{+}(S)$.

To illustrate the point further, consider the following example. Let $S = [0, 1]$ be equipped with the Lebesgue σ-algebra, and let C be a two-point set, $C = \{c_1, c_2\}$. A mapping (see A.5.1) $c_i \longmapsto \mu_i \in \Delta(S)$ defines a mapping $\kappa : F^{+}([0, 1]) \to \mathbb{R}_{+}^{C}$ in a way that each bet f on $[0, 1]$ is associated to the function $\phi \in \mathbb{R}_{+}^{C}$ which is the two-coordinate vector $\phi = (\int_{[0,1]} f d\mu_1, \int_{[0,1]} f d\mu_2)$. Clearly, κ is affine.

Suppose that C is equipped with the trivial algebra $\{\emptyset, C\}$. Then, while the set of real-valued mappings on C is isomorphic (in a set-theoretic sense) to the plane, the set of measurable functions on C consists of the constants only. If $\mu_1 \neq \mu_2$, then it is an easy matter to check that uncountably many measurable functions on $[0, 1]$ are associated to vectors whose first coordinate differs from the second, that is to non-constant functions on C. Since these are not measurable, we see that $range(\kappa) \not\subseteq F^{+}(C)$.

What drives the example is that while C is a two-point set it behaves, from the viewpoint of the measurable properties, as a one-point space. This property is only seemingly artificial. In fact, many natural quotient spaces display it. We will see examples of this sort when, in the second part, we study the effect of different information structures on the decision maker's behavior.

5.3 More models

Here, we want to push the above line of reasoning further. We want to allow for the possibility that our decision maker uses more than one model to solve the problem at hand, that of ordering the bets in $F^{+}(S)$. For instance, when it comes to forecasting the weather in Siberia, the decision maker might use, not only his knowledge of the weather in California, but also his knowledge of the weather in Nevada as well as that of any relation between the weather in California and Nevada, and so forth. A bit more formally, let M be the set of models for our decision maker. Each model $m \in M$ is a pair $(F^{+}(C_m), \succsim_{C_m})$, where $(C_m, \mathcal{C}_m)$ is some measurable space, and there is given a kernel $\kappa_m : F^{+}(S) \to F^{+}(C_m)$. Each model produces, by means of the associated kernel, an ordering of the bets in $F^{+}(S)$, and, generally speaking, different models produce different orderings. Then, the decision maker uses this collection of orderings to come up, if possible, with a solution for his problem. Just like before, the decision maker's understanding of M is described by two ingredients: (a) the set of bets on M that the he knows how to order; and (b) the way he orders them. The first is described by a σ-field, $\mathcal{M}$, of subsets of M or, equivalently, by the set $F^{+}(M)$ of nonnegative, bounded $\mathcal{M}$-measurable functions. The second, by a binary relation, $\succsim_M$, on $F^{+}(M)$. Intuitively, one might think of the latter as representing the decision maker's assessments of which model is more likely to apply.

To simplify the exposition, let us assume that, for each $m \in M$, the ordering $\succsim_{C_m}$ is represented by a functional $j_m : F^+(C_m) \to \mathbb{R}_+$. Such an assumption could be easily dispensed with, but at the cost of introducing a fairly cumbersome notation. Since here we are concerned more with the basic ideas rather than with formalities, the simplification is, in fact, harmless.

Following the idea outlined above, each function $f \in F^+(S)$ is now associated to the function $\phi_f \in \mathbb{R}_+^M$ which at point $m \in M$ takes the value

$$\phi_f(m) = j_m(f)$$

In a sense, the function ϕ_f is a description of the bet f when all models are taken into account.

Let $\kappa : F^+(S) \to \mathbb{R}_+^M$ be the function defined by $f \longmapsto \phi_f$. At this point, just like in 5.2, we have two possibilities. Either $range(\kappa) \subset F^+(M)$ or the inclusion does not hold. In the first case, the decision maker solves his problem by setting

$$f \succsim g \qquad iff \qquad \kappa(f) \succsim_M \kappa(g)$$

In the second, he concludes that his understanding of (his information about) M is not enough to solve the problem at hand, and must rely on other considerations.[7]

On the other hand, if it happens that $range(\kappa) \subset F^+(M)$ then the problem of ranking the bets in $F^+(S)$ splits into two parts. First, the mapping κ takes f into $\kappa(f)$, then $\kappa(f)$ is ranked by means of $\succsim_M$. If to fix ideas, we assume that $\succsim_M$ be itself represented by a functional $V : F^+(M) \to \mathbb{R}_+$, then we can summarize the discussion by means of the following diagram

$$\begin{array}{ccccc} F^+(S) & & \stackrel{\kappa}{\longrightarrow} & & F^+(M) \\ & I & \searrow & & \downarrow V \\ & & & & \mathbb{R}_+ \end{array}$$

that is

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

where $I : F^+(S) \to \mathbb{R}_+$ is defined by $I = V \circ \kappa$.

We conclude the section with a couple of remarks about the process of decision making that emerges from our construction. First, we suggest another possible interpretation of this process, which exploits the fact (noted in Section 4) that Case-Based Reasoning is a special case of Reasoning by Analogy. According to this point of view, each element in $m \in M$ can be thought of as a

[7] The last sentence is deliberately vague. Roughly, it means that the decision maker has to rely on considerations which are "nonmeasurable" with respect to his information about M. We study this problem in part III.

case or, more generally, as a collection of cases on the same domain. Then, the associated ordering on $F^+(S)$ can be thought of as the ordering suggested by the "experience" m and by the mapping κ_m. The set M is the collection of all such experiences and the field $\mathcal{M}$ describes the decision maker's view of how all these experiences (and the associated orderings on $F^+(S)$) fit together.

We would like to emphasize, however, that in general M may contain, not only the cases actually experienced by the decision maker, but also hypothetical cases and even cases the decision maker has never thought of. As a matter of fact, the field $\mathcal{M}$ and the functional V determine which cases in M affect the decision maker's ordering of the acts. For instance, suppose that $A \subset M$ is a subset of cases that the decision maker has never thought of and that, as such, are not going to affect his behavior. Then, with V defined by means of a probability measure P on $\mathcal{M}$, the condition $P(A) = 0$ would convey that such cases play no role.[8]

6 Choice-based foundations for models and analogies

The whole discussion above was heuristic. Some readers might find our definitions of model and analogy reasonable, others might be sceptical about them. The same can be said about the model of decision making which we outlined at the end of the previous section. It is a fact, nonetheless, that such a model admits a rigorous, choice-based, foundation.

Let $\mathcal{F}_0$ denote the set of simple Σ-measurable acts and $\mathcal{F}_c$ that of constant acts. Let $\succsim$ be a preference relation on $\mathcal{F}_0$ satisfying the following axioms.

A1 $\succsim$ is complete and transitive.

A2 (C-independence) For all $f, g \in \mathcal{F}_0$ and $h \in \mathcal{F}_c$ 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 $f, g, h \in \mathcal{F}_0$, 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 $f, g \in \mathcal{F}_0$, $f(s) \succsim g(s) \Longrightarrow f \succsim g$.

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

Recently, Ghirardato, Maccheroni and Marinacci [12] have shown that such preferences are completely described by the functional $I : F^+(S) \to \mathbb{R}$ defined by

$$I(u \circ f) = \alpha(f) \min_{P \in \mathcal{C}} \int_S u \circ f dP + (1-\alpha(f)) \max_{P \in \mathcal{C}} \int_S u \circ f dP$$

where $\mathcal{C}$ is a weak*-compact set of probability measures, $\alpha : F^+(S) \to [0,1]$ and u is a utility function on the prize space. It is readily seen that as special cases,

[8] The converse to this statement is not true. $P(A) = 0$ does not mean that the decision maker did not think of cases in A.

one obtains α-maxmin expected utility ($\forall \alpha \in [0,1]$), Choquet expected utility and Subjective expected utility.

With less emphasis on the form of the functional I and more on the structure that emerges from the theorem, we can reformulate the result as follows (the proof is in Appendix A.6).

Theorem 2 *Given a preference relation satisfying Axioms 1 to 5 there exist*
(i) a measurable space $(M, \mathcal{M})$
(ii) an affine mapping $\kappa : F^+(S) \to F^+(M)$
(iii) a functional $V : F^+(M) \to \mathbb{R}$
such that

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

Moreover, one can take $M = \mathcal{C} \subset \Delta(S)$ *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*

$$\phi_f(\mu) = \int\limits_S f d\mu$$

Remark 3 *A utility function on the prize space,* $u : X \to \mathbb{R}_+$*, produces a mapping from the set of acts into* $F^+(S)$ *given by* $f \longmapsto u \circ f$*. In the statement of the theorem, we have identified an act with its image in* $F^+(S)$*. This is harmless and simplifies the notation. We will stick with this convention throughout the paper.*

The theorem says that such a general model of decision making displays the same structure as the model we proposed at end of last section. In light of this observation, those discussions now provide a possible interpretation for the model emerging from Theorem 2. In other words, we can think of elements in M as "models" for the decision maker and of the functional I as consisting of two parts as in the diagram below

$$\begin{array}{ccc} F^+(S) & \xrightarrow{\kappa} & F^+(M, \mathcal{M}) \\ & I \quad \searrow & \downarrow V \\ & & \mathbb{R}_+ \end{array}$$

The first part is the mapping κ, which takes a bet $f \in F^+(S)$ into a function $\phi_f \in F^+(M)$. At model $m \in M$, ϕ_f takes a real value, $\phi_f(m)$, which can be interpreted as the evaluation of bet f corresponding to that model. All such possible evaluations are, then, collected together and lead to a single evaluation of the bet f by means of the functional V.

7 More general preferences

In this section, we are going to show that theorems of the above type hold for a much wider class of preferences. In our view, axioms A1 and A4 are very mild assumptions (at least from a positive viewpoint). Axiom A5 serves only to make things interesting. For this reason, we are going to provide representation theorems that, with respect to the previous one, drop either axiom A2 or axiom A3 or both. In Theorem 4 below, we drop only axiom A2 (C-independence) while maintaining the other axioms. This case is of special interest since it includes the class of Variational Preferences (Maccheroni, Marinacci and Rustichini [26]) which appears in important applications in Macroeconomics and in Finance (Hansen and Sargent [21]). From a more theoretical perspective, Ghirardato, Maccheroni and Marinacci [13] have shown that C-independence is the crucial property guaranteeing the complete separation between utility and "beliefs" (see [13] with regard to this terminology). Because of this, preferences satisfying A1 to A5 enjoy special properties, and are termed *invariant biseparable preferences*. Then, our Theorem 4 shows that the properties of the previous section hold beyond such a class. As a matter of fact, the combination of Theorem 4 and Theorem 2 show that they hold for all *biseparable preferences* (not necessarily invariant) and beyond that. The latter property appears, perhaps more clearly, in Corollary 5, which further extends the validity of Theorem 4. Finally, we show that our conclusions extend beyond the class of Archimedean preferences. By this, we mean that the property in axiom A3 fails not only for general acts, but for constant acts too. In such a case, there exists no real-valued utility on the prize space.

We begin by dropping the axiom of C-independence.

Theorem 4 *Let $\succsim$ be a preference relation on $\mathcal{F}_0$. If $\succsim$ satisfies A1 and A3 to A5, then there exist*
(i) a measurable space $(M, \mathcal{M})$
(ii) an affine mapping $\kappa : F^+(S) \to F^+(M)$
(iii) a functional $V : F^+(M) \to \mathbb{R}$
such that

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

Moreover, one can take $M = \Delta(S)$ and κ is defined by $f \longmapsto \phi_f$, where ϕ_f is the function that at point $\mu \in \Delta(S)$ takes the value

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

The proof is in Appendix A. 7. Just as in the proof of Theorem 2, we use the existence of a utility function on the prize space. Axioms A1 and A3 to A5 restricted to constants still guarantee the existence of a utility. However, because we have dropped A2, the utility is not necessarily linear. Obviously, the latter property might be restored by requiring that A2 hold for constants only (axiom LC, below). The same observation applies to the corollary below.

The next corollary further extends the validity of the previous theorem. In addition, it makes it clear that the class of preferences it applies to is wider than the class of *biseparable preferences* of Ghirardato and Marinacci [15]. The following axioms were introduced in [15], Section 3.

CE (Certainty Equivalents) For all $f \in \mathcal{F}_0$, $\exists x_f \in X : f \sim x_f$.

AAC (Archimedean Axiom for constants) For all $x, y, z \in X$, if $x \succ y \succ z$ then $\exists \alpha, \beta \in (0,1)$ such that $\alpha x + (1-\alpha)z \succ y \succ \beta x + (1-\beta)z$.

Corollary 5 *Let $\succsim$ be a preference relation on $\mathcal{F}_0$. If $\succsim$ satisfies A1, A4, A5, CE and ACC, then the same statement as in the previous theorem holds.*

Notice that the class of preferences dealt with in the corollary is wider than that in Theorem 4. In fact, axiom AAC is weaker than A3 in that it refers to constant acts only, and axioms A3 and A4 imply axiom CE but the converse is not true (see [15], Section 3). We also observe that, as a special case, the same statement holds if we replace CE with the continuity assumption S1 in [15], Section 3.3. For, in such a case, Lemma 29 in [15] implies that axiom CE is automatically satisfied. Such preferences are not necessarily biseparable in that they do not have to satisfy the assumption of Binary Comonotonic Independence in [15], Section 3.3.

We now move to the non-Archimedean case. We first introduce the following axiom, which weakens A2 in that it requires that independence holds for elements in the prize space only.

LC For all $x, y, z \in X$ and for all $\alpha \in (0,1)$

$$x \succ y \qquad \Longleftrightarrow \qquad \alpha x + (1-\alpha)z \succ \alpha y + (1-\alpha)z$$

Given axiom LC, the extension of the above results to the non-Archimedean case obtains along similar lines. In fact, axioms A1, A4 and LC guarantee the existence of a utility representing the ordering on the prize space. Since we gave up the Archimedean axiom, such a utility is not valued in the reals but rather in an ordered vector space OV^* (see Hausner [22]). Without loss, we can assume that the latter is a Banach space. Hence, acts can be identified to mappings from S to this Banach space. By axiom CE, we can define an operator from such mappings to OV^*, which represents the decision maker's ordering of the acts. Finally, we define κ and V just like we did in the previous theorems. The only difference is that the integrals of the previous theorems as well as V are now OV^*-valued (the integrals are, in fact, Bochner integrals). Summarizing, we have

Theorem 6 *Let $\succsim$ be a preference relation on $\mathcal{F}_0$. If $\succsim$ satisfies A1, A4, A5, CE and LC, then there exist*
(o) a utility function $u : X \to OV^*$, where OV^* is a Banach space.
(i) a measurable space $(M, \mathcal{M})$
(ii) an affine mapping $\kappa : F_0(S, OV^*) \to F(M, OV^*)$ [9]

[9] $F_0(S, OV^*)$ is the set of simple measurable mappings $S \to OV^*$ and $F(M, OV^*)$ that of measurable mappings $M \to OV^*$.

(iii) a mapping $V : F(M, OV^*) \to OV^*$
such that

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

Moreover, one can take $M = \Delta(S)$ *and* κ *is defined by* $f \longmapsto \phi_f$, *with* ϕ_f *being the function that at point* $\mu \in \Delta(S)$ *takes the value*

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

where the integral on the RHS is a Bochner integral.

8 Comments on the representation

The structure which emerges from the representation theorems contains several ingredients. A closer look at it is, now, in order. The theorems state that we can think of the decision process as consisting of two stages: first, the decision maker maps the bets $F^+(S)$ into $F^+(M)$ by means of κ, then he orders $F^+(M)$, hence $F^+(S)$, by means of V. To begin, we must point out that it is not observable whether or not the decision process actually occurs in this way. Pictorially, what is observable is only the diagonal arrow in the diagram at the end of Section 6. Yet, as we have noted already, the content of the above theorems is precisely that no choice-based theory can dismiss this possibility. In other words, we can always think of the decision process as consisting of the two stages described above. This brings us to an interesting issue. If we must admit that the decision process involves two sets, S and M, then we must admit that there are two sources of uncertainty as well as two potential sources of information (about S and about M). We comment on this in 8.2 below. Before that, however, it is useful to spend a few words about the nature of the space M. Since we have already argued that the decision maker can be uncertain about M, it would be good to know what this uncertainty is about.

8.1 M is a set of risky descriptions

A notable feature emerging from the theorems of the previous sections is that the set of models for the decision maker can always be described by means of the set of all probability measures on S. In this regard, the reader should notice that while this result appears in Theorem 2 as a consequence of the representation theorem given in [12], the same result in Theorems 4 and 6 is independent of that representation. Of course, the representation of M as a set of probability measures on S is by no means unique. Any set that is measurably isomorphic with it will do as well. Yet, the very possibility of describing it by means of such probabilities is of special significance for a theory of decision making. It states that M can always be thought of as the collection of all probabilistic descriptions associated with the domain of the acts. There is more to it, however. An additional insight comes from the observation that the mapping κ, which links S and M in the theorems, 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 of model m. In the spirit of the Knightian distinction between risk and ambiguity, we can then think of M as the collection of all risky descriptions of the problem faced by the decision maker.

8.2 Two types of information

No matter how one interprets M, it is a set that is linked to S by a certain relation. Moreover, the decision maker takes this relation into account when it comes to ordering the bets in $F^+(S)$.

Generally speaking, the decision maker is uncertain about both S and M.[10] In part II, we will argue that these are uncertainties of inherently different types. For now, we want to limit ourselves to a simpler observation. 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 $F^+(S)$, because M and S are linked to each other. For instance, in our example above, the decision maker's forecast about the weather in Siberia is going to be affected by any information that he obtains about the weather in California or about the relation between the weather in California and that in Nevada. We shall see (part II) that the study of how information about M translates into information about S is a nontrivial issue. For future reference, it is useful to summarize the content of this discussion in the following schematic way

Two sets → Two types of uncertainty → Two types of information

8.3 The space M. The normative use of the model

M emerges from the theorems not only as a set, but also as a space, that is a set equipped with a certain structure (the Borel tribe described in the theorems). Such a structure represents the decision maker's understanding of the collection of risky descriptions associated with S. This is a crucial ingredient. Ultimately, the decision maker has to fit together all those descriptions, and combine them so to generate a unique ordering of the acts. It is, then, intuitively clear that the decision maker's understanding of M plays a crucial role in determining such an outcome. This brings us to our next observation, which regards the ability of using the model emerging from Theorems 2, 4 and 6 in a normative way.

[10] It is interesting to observe that the extreme case of "certainty about the model" implies SEU theory. In fact, the case of certainty about the model corresponds to the kernel defined by the constant mapping $m \longmapsto \mu \in \Delta(S)$, $\forall m \in M$. Hence, for $f \in F^+(S)$, we have $f \longmapsto \phi_f(m) = \int f d\mu$, $\forall m \in M$. Therefore, any V compatible with the representation theorems implies $V \circ k(f) = \int f d\mu$. As we shall see, this is not the only case compatible with SEU.

If we are going to use the model in a normative way, then the algebra on M represents (by assumption) the decision maker's understanding of the set of all risky situations. This raises an important issue. To see it, let us go back to the representation theorems. In those theorems, M emerges equipped with the Borel tribe generated by the weak*-topology. This is precisely the algebra that guarantees that the mapping κ is measurable in the sense of Appendix A.5.1. Or, if we go back to our discussion about Reasoning by Analogy, this is what guarantees that the decision maker's understanding of the set M is enough to order the bets in $F^+(S)$. From this viewpoint, the theorems should be interpreted as saying that if the decision maker's information on M is described by the Borel tribe generated by the weak*-topology on M, then the bets in $F^+(S)$ are ordered according to the functional $V \circ \kappa$. Immediately, one notices that the condition about the measurable structure of M is, in fact, an informational requirement. That is, it states that the decision maker's information about M must be of a certain type. One obvious question is, how demanding is such a requirement?. A look at the properties of the Borel tribe generated by the weak*-topology tells us that this is not a light condition. For instance, it implies that, on the basis of his information, the decision maker is able to distinguish between any two models. In other words, given any two models, m_1 and m_2, the decision maker can always construct (by using his information only) a statistical test which allows him to reject/accept the hypothesis that the true model is m_1 and not m_2.

In part II, we will show that this is precisely the necessary condition allowing a (Bayesian) decision maker to summarize both his uncertainty about M and his uncertainty about S by means of a single probability measure on S. We will also show that this condition is not satisfied in many instances. In particular, we will show that it is not so in Ellsberg's experiments.

9 A Borel Setting

This section is a (very) brief digression. Here, I introduce some additional assumptions guaranteeing, among other things, that all the priors in the model are countably additive. While these assumptions play a role only in the sequel, it is economically convenient to introduce them following the detailed exposition of the other axioms contained in Sections 6 and 7. In this way, the reader would have a clear picture of the model that I am going to be using in the remaining part of the work.

From now on, we are going to focus on preferences satisfying axioms A1 to A5. This choice allows us to avoid a certain amount of tedious qualifications and cumbersome notations. Nevertheless, as the reader can readily check, all the theorems that we state in the sequel extend with minor qualifications to the more general classes of preferences dealt with in Section 7. We also introduce three additional assumptions which we maintain throughout the work, unless otherwise stated.[11] The first two are about the measurable space (S, Σ). We

[11] A notable exception is the section devoted to the case of a finite domain of the acts (see

assume that S has the cardinality of the continuum and that (S, Σ) is a standard Borel space (see A.9 for a definition). The Isomorphism Theorem for Borel spaces (see A.9) implies that, without any loss, the reader can think of (S, Σ) as $[0, 1]$ equipped with the Lebesgue σ-algebra Λ. The third assumption is on preferences, and naturally complements the previous two. We assume that preferences satisfy the Axiom of Monotone Continuity (see A.9). In this case, all the measures appearing in Theorem 2 are countably additive (see A.9). For a discussion about this axiom the reader is referred to [2], where the Axiom was originally introduced, and to [34].

We have two main motivations for introducing these assumptions. One is that we want to re-examine Savage theory in light of the study of Information we will pursue in part II. Savage theory demands that the set S be infinite, which explains, at least in part, the assumption about the cardinality of S. The second and more important perhaps, is that the study of the uncountable standard Borel case leads, as we shall see, to a deeper understanding of the finite case. In this regard, the reader should remind him/herself that a finite set equipped with the power set is a standard Borel space. Of course, when moving the uncountable case, one cannot retain both the power set and the property of being standard Borel, and it is a debatable matter, to say the least, which of these two representations is the appropriate extension of the finite case. In fact, there are very good and well-known reasons to prefer the latter. We will examine this problem in part II.

We conclude this discussion by observing that under the new assumptions the mapping κ of Theorem 2 is, in fact, a kernel (see A.9).

10 An Illustration: Ellsberg's experiment

Not surprisingly, Ellsberg's three-color urn experiment is the perfect example to illustrate the concepts we have developed so far. In fact, the ingredients of the model emerging from Theorem 2 and the issues related to those 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.[12] Finally, $F^+(M)$ (considered as a set) can

part II).

[12] As the reader might observe, this description could be enlarged by adding all the possible permutations of the 90 balls. This is legitimate as, in principle, one could order the balls, and give the decision maker information about such an ordering (for instance, "the 29th ball is

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. Moreover, the existence of a relevant connection between the two system of bets is a subjective statement, that is, it pertains to the decision maker himself.

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 would like to 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 $F^+(S)$. 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 the 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: $\mathcal{M}$, the σ-field on the set of models. Here, the question is the following: in the three-color urn experiment, the information, "the number of balls that are either blue or green is 60", is the only information about the possible configurations of the urn available to the decision maker. With M being the set of all such possible configurations, what σ-field on M describes this information? We will have to introduce a number of additional concepts before attacking this problem, and we will do so only in part II. For now, we want to conclude our illustration by giving a continuous version of the three-color experiment, which we use later in the work. In its continuous version, the urn is identified to the interval $[0, 1]$, which we should think of as partitioned into three subsets, labeled R, B and G. The set of bets is

$$F^+([0,1]) = \{f \mid f : [0,1] \to \mathbb{R}_+, f \text{ bounded and } \Lambda\text{-measurable}\}$$

where Λ is the Lebesgue σ-algebra. The set of models, M, has the same meaning as before, and is described by the set of probability measures on $([0,1], \Lambda)$. The decision maker's information is about this set and not about the domain of the bets. It takes the form, "the true model belongs to the subset $\Omega \subset \Delta([0,1])$ such that $\mu(R) = 1/3$ for every $\mu \in \Omega$". This still begs the question of what is the σ-field on the set of models which describes this information.

11 The Bayesian view

Having noticed that Ellsberg's experiment fits the model of Theorem 2 perfectly, it makes sense to examine the Bayesian argument stating that, even in our context, the decision maker's choices should conform to an expected utility criterion.

red"). One can account for this possibility by proving a representation theorem like the one above, but where the set $\Delta(S)$ is replaced by the set of automorphisms of the algebra Σ. We will do so in part II when we study Ellsberg's Paradox.

Let us recall that the crucial features of the experiment are that (i) the decision maker is uncertain about two objects, the domain of the bets and the set of models; and (ii) he obtains information about the set of models and not about the domain of the bets.

To begin, a Bayesian would argue that since the decision maker is uncertain about the set of models, then such an uncertainty should be described by means of a probability measure, P, on that set. This is tantamount to saying that the decision maker obeys SEU for the system of hypothetical bets $F^+(M)$. The set of models is a set of probability measures on S. Hence, the decision maker would be described by a measure over those measures.

The next step is to argue that, when it comes to ordering the bets in $F^+(S)$, the decision maker will compute the average of those measures using the "weights" given by P. By doing so, he will end up with a single measure on S, which describes his uncertainty. This is an old story. If a Bayesian decision maker is described by means of a measure P over a set of measures on the domain of the acts, then he is equivalent, from an observational viewpoint, to a Bayesian decision maker described by a single measure on the domain of the acts. This measure is obtained by averaging out those measures, the weights being given by P.

Assuming its validity, this reasoning has an important corollary, namely, information about the set of models always translates into information about the domain of the bets. In fact, if the Bayesian decision maker obtains some information about the set of models, he will use it to compute a system of conditional probabilities on that set, one for each element of the partition determined by the information. Once this is done, he will apply the above reasoning, thus obtaining, once again, a single measure on S.

While this type of reasoning is very popular, the conditions that guarantee its validity are far from being trivially satisfied. In fact, the validity of this reasoning is subject precisely to that informational requirement mentioned in 8.3. In part II, we study such conditions and show that, whenever they are not satisfied, "rational" decision makers cannot conform to the choices postulated by the SEU paradigm. Moreover, we will argue that the Ellsberg's choices might rationally emerge as a consequence of this fact.

12 Conclusion

In this paper, we have shown that virtually any axiomatic model of decision making under uncertainty displays a certain structure. Such a structure involves two sets – the set of models for the decision maker and the domain of the acts – and allows us to think of the decision process as occurring in two stages. We showed that this procedure can be thought of as describing a general form of analogical reasoning.

The appearance, in any given choice problem, of the two different sets implies that the decision maker is uncertain about two different objects and that he may receive information about any of them. This raises the issue of how information about the set models affects the decision maker's ranking of the available alternatives. This is the topic we study in the second part.

There, we focus on decision makers who satisfy the SEU paradigm if their information on the set of models is sufficiently fine. We then show, that the same decision makers fail to obey the SEU paradigm in correspondence of coarser information structures. We explicitly exhibit two such information structures, and argue that one of them is naturally associated with Ellsberg's three-color urn experiment. We also use the properties of such information structures to give a formal definition of Knightian Uncertainty.

References

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

[2] Arrow K.J. (1974), Exposition of a theory of choice under uncertainty, in Essays in the Theory of Risk-Bearing, North Holland.

[3] Chalmers D.J., R.M. French and D.R. Hofstadter (1992), High-level perception, representation, and analogy: A critique of AI methodology", *Journal of Experimental & Theoretical Artificial Intelligence* **4**, 185-211.

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

[5] Diestel J. and J. J. Uhl (1977) Vector Measures, AMS Mathematical Surveys No. 15.

[6] Dictionary of Philosophy of Mind, www.artsci.wustl.edu/~philos/MindDict/analogy.html

[7] Ellsberg, D. (1961), "Risk, Ambiguity and the Savage Axioms", *Quarterly Journal of Economics* 75, 643-69.

[8] Epstein L.G. and M. Schneider (2003), Recursive multiple-priors, *Journal of Economic Theory* **113**, 1-31.

[9] Epstein L.G. and J. Zhang (2001), Subjective probabilities on subjectively unambiguous events, *Econometrica* **69**, 265-306.

[10] Falkenheimer B., K.D. Forbus and D. Gentner (1989), The Structure mapping engine: algorithm and examples, *Artificial Intelligence* **41**, 1-63.

[11] Forbus K.D., D. Gentner, A.B. Markman and R.N. Ferguson (1998), Analogy just looks like high level perception: Why a domain-general approach to analogical mapping is right, *Journal of Experimental & Theoretical Artificial Intelligence* **10:2**, 231-57.

[12] Ghirardato P., F. Maccheroni and M. Marinacci (2003), Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, forthcoming.

[13] Ghirardato P., F. Maccheroni and M. Marinacci (2003), Certainty Independence and the Separation of Utility and Beliefs, *Journal of Economic Theory*, forthcoming.

[14] 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.

[15] Ghirardato P. and M. Marinacci (2001), Risk, ambiguity and the separation of utility and beliefs, *Mathematics of Operation Research* **26**, 864-90.

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

[17] Gilboa I. and D. Schmeidler (2001), A Theory of Case-Based Decisions, Cambridge University Press.

[18] Gilboa I. and D. Schmeidler (2003), Inductive Inference: An Axiomatic Approach, *Econometrica* **71**, 1-26.

[19] Gul F. and W. Pesendorfer (2001), Temptation and Self-Control, *Econometrica* **69**, 1403-35.

[20] Halmos (1963), Lectures on Boolean Algebras, Van Nostrand.

[21] Hansen L. and T. Sargent (2000), Wanting robustness in macroeconomics, mimeo.

[22] Hausner M. (1954), Multidimensional Utilities, in Decision Process, R.M. Thrall, C.H. Coombs and R.L. Davis eds, John Wiley.

[23] Hausner M. and J.G. Wendel (1952), Ordered vector spaces, *Proceedings of the AMS* **3**, 977-82.

[24] Karni E. (2004), Subjective Expected Utility Theory without States of the World, The Johns Hopkins University, miemo.

[25] Kastler D. (1982), On A. Connes' Noncommutative Integration Theory, *Communications in Mathematical Physics*, **85**.

[26] F. Maccheroni, M. Marinacci and A. Rustichini (2004), Variational representation of preferences under ambiguity, ICER WP 5/04

[27] Marinacci M. (1997), A simple proof of a basic result for multiple priors, University of Toronto, mimeo.

[28] Morrison C.T. and E. Dietrich (1995), Structure-mapping vs. high-level perception: the mistaken flight over the explanation of Analogy", *Proc. 17th Annual Conference of Cognitive Science Society*, 678-82.

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

[30] Kahneman D. and A. Tversky (1979), Prospect Theory: An Analysis of Decision under Risk, *Econometrica* **47**, 263–91.

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

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

[33] Srivastava S.M. (1998), A Course on Borel Sets, Springer.

[34] Stinchcombe, Maxwell (1997), Countably additive subjective probabilities. *Review of Economic Studies* **64**, 125-46.

[35] Stinchcombe, Maxwell (2003), Choice with Ambiguity as Sets of Probabilities, mimeo.

APPENDIX

A.5.1 Kernels

The concept of kernel is the building block of our definition of reasoning by analogy. In fact, kernels play an important role throughout this work. Thus, it is appropriate to gather here a few basic facts about them. For more, the reader should consult [25].

Let $(Y, \mathcal{Y})$ and $(Y', \mathcal{Y}')$ be two measurable spaces. A kernel from Y to Y' is a mapping $\nu : F^+(Y) \to F^+(Y')$ which is

(i) affine: $\nu(\alpha f + \beta g) = \alpha\nu(f) + \beta\nu(g), \qquad \alpha, \beta \in \mathbb{R}_+$

(ii) normal: $f_n \nearrow f \Longrightarrow \nu(f_n) \nearrow \nu(f), \qquad n \in \mathbb{N}$

In other words, a kernel is a representation (that is, a mapping which is structure-preserving) of the positive measurable functions on Y into the positive measurable functions on Y'. Any kernel can be equivalently described by means of a mapping, $Y' \to \Delta^+(Y)$, from Y' to the set, $\Delta(Y)$, of positive measures on Y. In fact, one associates the element $y' \in Y'$ to the measure $\nu^{y'} \in \Delta(Y)$, which is defined by the equation

$$\nu^{y'}(A) = (\nu(\chi_A))\,(y'), \qquad for \quad every \quad A \in \mathcal{Y}$$

and the mapping

$$Y' \to \Delta^+(Y) \qquad defined \quad by \qquad y' \longmapsto \nu^{y'} \tag{1}$$

is *measurable* in the sense that for every $A \in \mathcal{Y}$, the function on Y' which takes the value $\nu^{y'}(A)$ at the point $y' \in Y'$ is an element of $F^+(Y')$.

The converse of this is not necessarily true. While any mapping $Y' \to \Delta(Y)$ defines a mapping $\tilde{\nu} : F^+(Y) \to \mathbb{R}^{Y'}$, it is not guaranteed that $range(\tilde{\nu}) \subset$

$F^+(Y')$. That is, $\tilde{\nu}$ might fail to be measurable in the above sense. For instance, given $Y' \to \Delta(Y)$, we can define $\tilde{\nu}$ by

$$\tilde{\nu} : f \longmapsto \tilde{\nu}(f)$$

where $\tilde{\nu}(f)$ is the function in $\mathbb{R}^{Y'}$ which at point $y' \in Y'$ takes the value

$$\tilde{\nu}(f)(y') = \int_Y f d\nu^{y'} \tag{2}$$

and it is clear that whether or not $\tilde{\nu}(f)$ is a measurable function on Y', i.e. is an element of $F^+(Y')$, depends on the σ-algebra $\mathcal{Y}'$ that we have on Y'. On the other hand, it is immediate to check that if $range(\tilde{\nu}) \subset F^+(Y')$, then $\tilde{\nu}$ is affine and normal and, therefore, it is a kernel, and all kernels are defined essentially in this way (see equation (2)).

A.6 Theorem 2

Recall that S is the domain of the acts, Σ is a fixed field of events in S and X is the prize space. We assume that X is a convex set. As it is well-known, such an assumption can be justified by thinking of X as the set of lotteries on some given set of outcomes as in Anscombe and Aumann [1]. Alternatively, the assumption can be justified on the basis of the axiomatization of preferences given in Ghirardato, Maccheroni, Marinacci and Siniscalchi [14].

Let $\mathcal{F}_0$ be the set of simple Σ-measurable acts (acts that take only finitely many values in X) and let $\mathcal{F}_c$ be the set of constant acts. Let $\succsim$ be a binary relation on $\mathcal{F}_0$, and let $\succ$ denote its asymmetric part.

Elements in the prize space are identified to the set of constant acts. Axioms 1 to 4 in the text imply the existence of a linear utility on the prize space. As noticed (remark 3), this allows to identify acts with the set $B_0(S, W)$ of bounded Σ-measurable simple functions on S which take values in $W = range(u)$. Gilboa and Schmeidler [16] have shown that preferences satisfying A1 to A5 are represented by a functional $I : B_0(S, W) \to \mathbb{R}$ which is C-independent, positively homogeneous (which allows to extend by homogeneity I to the whole B_0 – the set of bounded Σ-measurable simple functions on S), monotone and supnorm continuous. Finally, the latter property allows to extend I to the set of bounded Σ-measurable functions on S as B_0 is norm dense in the latter set.

Proof of Theorem 2. In [12], it was shown that I takes the form

$$I(u \circ f) = \alpha(f) \min_{P \in \mathcal{C}} \int_S u \circ f dP + (1 - \alpha(f)) \max_{P \in \mathcal{C}} \int_S u \circ f dP$$

for $\mathcal{C}$ a weak*-compact set of probability measures, $\alpha : F^+(S) \to [0, 1]$ and u a utility function on the prize space.

Set $M = \mathcal{C} \subset \Delta(S)$, and let $\mathcal{M}$ be the Borel tribe generated by the weak*-topology on $\Delta(S)$. Then, let κ be defined by $f \longmapsto \kappa(f)$ where

$$\kappa(f)(\mu) = \int f d\mu \qquad , \qquad \mu \in \Delta(S)$$

Notice that $\mathcal{C}$ is measurable in $\Delta(S)$ because it is weak*-closed.

In order to prove the theorem, we need to show two things:

(i) κ is measurable in the sense of A.5.1, i.e. $range(\kappa) \subset F^{+}(\mathcal{C})$;

(ii) the function $\alpha : F^{+}(S) \to [0,1]$ is compatible with the nucleous of equivalence of κ, that is if $f, g \in F^{+}(S)$ are such that $\kappa(f) = \kappa(g)$, then $\alpha(f) = \alpha(g)$. This will allow us to define α on $F^{+}(\mathcal{C})$ rather than on $F^{+}(S)$.

To prove (i) observe that the function $\kappa(f) : \Delta(S) \to R_{+}$ is trivially continuous for the weak*-topology on $\Delta(S)$. Hence, it is measurable for the Borel tribe generated by that topology.

(ii) was already observed by [12]. We provide here a slightly different proof. Let $f, g \in F^{+}(S)$ be such that $\kappa(f) = \kappa(g)$. In the terminology of [12], this implies that f is unambiguously indifferent to g. Since unambiguous preference is a subrelation of the decision maker's preference relation over acts, this implies that f is indifferent to g, that is $I(f) = I(g)$. Since $\forall f \in F^{+}(S)$, $\alpha(f)$ is uniquely defined by

$$\alpha(f) = \frac{I(f) - \max\limits_{\mu \in \Delta(S)} \kappa(f)(\mu)}{\min\limits_{\mu \in \Delta(S)} \kappa(f)(\mu) - \max\limits_{\mu \in \Delta(S)} \kappa(f)(\mu)}$$

we have $\alpha(f) = \alpha(g)$. ■

A. 7 Theorems 4 to 6

Proof of Theorem 4. Axioms A1 and A3 to A5 restricted to constants imply the existence of a utility function (not necessarily linear) on the prize space. Axioms A3, A4 (see for instance Gilboa and Schmeidler [16], proof of Lemma 3.2) imply that for each $f \in \mathcal{F}_0$there exists $x_f \in X$ such that $f \sim x_f$. Define $J : F_0 \to \mathbb{R}$ by $J(f) = u(x_f)$. Clearly, J represents $\succsim$. Set $W = range(u)$. Let $B_0(\Sigma, W)$ denote the set of bounded, Σ-measurable simple functions with range in W. The utility function $u : X \to R$ defines an operator $T_u : F_0 \to B_0(\Sigma, W)$ by $T_u(f) = u \circ f$. Define $I : B_0(\Sigma, W) \to R$ as the unique operator that makes the diagram below commute

$$\begin{array}{ccc} F_0 & \xrightarrow{T_u} & B_0(\Sigma, W) \\ & J \searrow & \downarrow I \\ & & \mathbb{R} \end{array}$$

$B_0(\Sigma, W)$ is a subset of B_0, the set of bounded, Σ-measurable simple functions. B_0 equipped with the supnorm is Banach space. Hence, it has *sufficiently many* continuous linear functionals. That is, if $a, b \in B_0$, $a \neq b$, there exists a continuous linear functionals L on B_0 such that $L(a) \neq L(b)$. By Riesz representation theorem, a continuous linear functional on B_0 has the form $\int a d\mu$, with μ a finitely additive measure on Σ. Hence, the mapping $\kappa : B_0(\Sigma, W) \to F(\Delta(S))$ defined as in the statement of the proposition is one-to-one. As observed in the proof of the previous theorem, $\kappa(a)$ is a measurable function on $(M = \Delta(S), \mathcal{M})$

with $\mathcal{M}$ being the Borel tribe generated by the weak*-topology. Finally, define V as the unique functional that makes the following diagram commute

$$\begin{array}{ccc} B_0(\Sigma, W) & \overset{\kappa}{\longrightarrow} & F^+(M, \mathcal{M}) \\ & I \quad \searrow & \downarrow V \\ & & \mathbb{R} \end{array}$$

■

Proof of Corollary 5. Axioms A1, A4 and AAC imply the existence a a real-valued utility on the prize space. CE implies that for each $f \in \mathcal{F}_0$there exists $x_f \in X$ such that $f \sim x_f$. Define $J : F_0 \to \mathbb{R}$ by $J(f) = u(x_f)$. Clearly, J represents $\succsim$. Then, proceed as in the above proof. ■

Proof of Theorem 6. By A1, A4 restricted to constants and LC, the set X along with the order and the mixture operations is a (non-Archimedean) utility space in the sense on Hausner [22]. Hausner [22] and Hausner and Wendel [23] have shown that there exists an order-preserving embedding $\tilde{u}$ (a utility, not real-valued) of X into an ordered vector space OV. Denote by $\geq$ the order on OV.

Define a norm $|\cdot|$ on OV, and denote by OV^* its completion. That is, OV^* is a Banach space. By embedding OV into OV^*, each act $f \in \mathcal{F}_0$ is associated to the mapping $\tilde{u} \circ f$ from S to the Banach space OV^*. For each $f \in \mathcal{F}_0$, $\tilde{u} \circ f$ is measurable and, being a simple function, is strongly measurable. Let $F_0(S, OV^*)$ be the set of all measurable simple mappings $S \to OV^*$.

Endow $F_0(S, OV^*)$ with the norm defined by

$$\|f\| = \{\sup |f(s)| : s \in S\} \qquad , \qquad f \in F_0(S, OV^*)$$

Since OV^* is a Banach space so is $F_0(S, OV^*)$.

By CE (proceeding just like in the previous two proofs), there exists a mapping $J : \mathcal{F}_0 \to OV^*$ such that $f \succsim g$ iff $J(f) \geq J(g)$.

For $\mu \in \Delta(S)$ and $a \in F_0(S, OV^*)$, denote by $\int a d\mu$ the Bochner integral (see for instance [5]).

On $M = \Delta(S)$, we define a topology τ as the coarsest topology such that for any bounded strongly measurable function $b : S \to OV^*$

$$\mu_n \to \mu \qquad \Longrightarrow \qquad \int b d\mu_n \to \int b d\mu$$

Let $\mathcal{B}$ be the Borel tribe generated by τ and let $F(\Delta(S), OV^*)$ be the space of measurable mappings $M \to OV^*$ (these are the mappings that are measurable with respect to $\mathcal{B}$ and the Borel tribe generated by the norm topology on OV^*).

Define $\kappa : f \longmapsto \phi_f$ as in the statement of the theorem. The mapping is clearly injective [If $f, g \in F_0(S, OV^*)$ and $f \neq g$ then (since they are both simple) there exists $A \in \Sigma$ such that $f(s) = x \in OV^*$ and $g(s) = y \in OV^*$for any $s \in A$ and $x \neq y$. Pick $\mu \in \Delta(S)$ so that $\mu(A) = 1$. Then, $\int f d\mu \neq \int g d\mu$]. Moreover, for each $f \in F_0(S, OV^*)$ the mapping ϕ_f is continuous for the topology τ and, hence, measurable. Hence, $range(\kappa) \subset F(\Delta(S), OV^*)$. ■

A. 9 A Borel setting

A.9.1 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 [33], Theorem 3.3.13)

Theorem 7 (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.

In part II, we will often use the well-known fact that the reduced Borel measure algebra on $[0, 1]$ (Borel sets modulo Borel sets of measure zero) and the reduced Lebesgue measure algebra on $[0, 1]$ (Lebesgue sets modulo Lebesgue sets of measure zero) are the same (see [20], p.68).

A.9.2 Monotone Continuity

In [12], Ghirardato, Maccheroni and Marinacci introduced the following relation on $\mathcal{F}$, which they termed unambiguous preference relation.

Definition 8 ([12]) *Let $f, g \in \mathcal{F}$. f is unambiguously preferred to g, $f \succsim^* g$, if*

$$\lambda f + (1 - \lambda)h \succsim \lambda g + (1 - \lambda)h$$

for all $\lambda \in (0, 1)$ and $h \in F$.

As shown in [12] (Sec. B.3), for a preference relation satisfying A1 to A5, the following axiom is equivalent to the property that all the priors in the representation are countably additive.

A6 (Monotone Continuity) 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 introduction of Axiom 6 allows us to conclude that the mapping κ in Theorem 2 is a kernel. In fact, it is immediate that κ is affine. Without Axiom 6, this is all we can say. In particular, we cannot conclude that κ is normal because with finitely additive measure the Dominated Convergence Theorem need not hold. However, with Axiom 6, all the measures in the theorem are countably additive. Hence, since every f is bounded, it is immediate to verify that κ is normal.