Partially-specified probabilities: decisions and games

Ehud Lehrer*

August 31, 2011

Abstract:

The paper develops a theory of decision making based on partially-specified probabilities. It takes an axiomatic approach using Anscombe-Aumann's setting, and is based on the concave integral for capacities. This theory is then expanded to interactive models in order to extend Nash equilibrium by introducing the concept of *partially-specified equilibrium*.

Keywords: Fat-free act; strongly fat-free act; partially-specified probability; decision making; ambiguity aversion; partially-specified equilibrium

---

*The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. e-mail: lehrer@post.tau.ac.il
The author acknowledges Yaron Azrieli, David Schmeidler, Uzi Segal and Peter Wakker for their helpful suggestions. The anonymous referee of AEJMicro contributed a great deal to the quality of the paper. Roee Teper is specially acknowledged for his help in revising it. The author acknowledges the support of the Israel Science Foundation through Grant 538/11.

In Ellsberg paradox, a decision maker does not obtain a full information about the real distribution of the balls' colors. Instead, she is informed that out of 90 balls that might be white, black or red, 30 are red. She does not know how many balls are white, for instance. Based on this partial information she is asked to choose between various gambles.

In this scenario, the decision maker first obtains partial information about the frequency of colors, and only then makes up her mind. However, a modeler who observes only her choices could infer that she behaves as if prior to making a decision, she obtained some information about the underlying frequency of colors.

Inspired by Ellsberg paradox, this paper develops a theory of decision making under uncertainty that is based on *partially-specified probabilities*. The first part of the paper joins a prominent list of papers,[1] among which are von-Neumann and Morgenstern (1944), Savage (1954), Anscombe and Aumann (1963), Schmiedler (1989), Gilboa and Schmeidler (1989), and Maccheroni et al. (2006), that take an axiomatic approach. Like these papers, this one presents a list of axioms that have a behavioral appeal, and represents it by a decision making model.

The subject matter of the first part are preference orders over Anscombe-Aumann's acts (Anscombe and Aumann, 1963) that assign each state of nature a lottery. Preference orders that satisfy four rather standard axioms and an additional minor one are shown to behave as if based on partially-specified probabilities. The four axioms are completeness and transitivity, continuity, independence and monotonicity. The first three appear (with a minor variation) in von-Neumann and Morgenstern (1944), while monotonicity has been introduced by Anscombe and Aumann (1963). The main difference between the current axiomatization and previous axiomatizations that are based on Anscombe-Aumann's setting is in the formulation of the independence axiom.

A *fat-free* act is characterized by the property that if at least one of the lotteries assigned to states is replaced by a worse lottery, the resulting act is strictly inferior to the original one. Figuratively speaking, in a fat-free act, no fat can be cut without affecting its quality. In contrast, in a fat-rich act there is at least one state whose assigned lottery can be downgraded without affecting its quality, thus obtaining a modified act which is equivalent to the original one. In other words, fat comes in a form of excessively high lotteries assigned to one or a few states. These lotteries can be downgraded while leaving the quality of the act unharmed. A *strongly fat-free* act is one that remains fat-free even when mixed with a constant act.

By definition, fat embedded in an act does not contribute to its quality. This does not

[1] An exhaustive list would require a few more pages.

mean, however, that excessively high lotteries – 'fat' ones – play no role. Excessively high lottery embedded in a fat-rich act might react synergetically with another act. That is, the fat contained in one act might enhance the quality of a mixture with another act to a level higher than that of each separately. This cannot happen when an act is mixed with a strongly fat-free act. No matter how excessively high the lotteries of an act are, when combined with a strongly fat-free act, the quality of the outcome cannot be higher than that of each.[2]

The independence axiom used here applies to strongly fat-free acts. It states that when one act is preferred to another, this preference remains intact even when the two acts are mixed with a strongly fat-free act. This axiom, together with the axioms of completeness, transitivity, continuity, monotonicity and an additional minor axiom (which states that a constant act is strongly fat-free), implies that the preference order is determined by evaluating acts based on partial information obtained about the underlying probability.

The particular independence axiom employed here is weaker than that of Anscombe and Aumann (1963) and stronger than that of Gilboa and Schmeidler (1989). The independence axiom of Anscombe and Aumann (1963) applies to all acts, while Gilboa and Schmeidler (1989) assume only certainty independence that applies to constant acts alone.

It turns out that a preference order that is based on a partially-specified probability satisfies the axiom of uncertainty aversion, introduced by Gilboa and Schmeidler (1989). This axiom states that if two acts are preferred to a third, then any mixture of them is still preferred to that third act. Thus, strengthening Gilboa and Schmeidler's independence axiom enables one to drop their uncertainty aversion axiom and still get a model that satisfies it.

The decision maker whose preference order satisfies the aforementioned axioms looks as if she knows the probability (either subjective or objective) of some, typically not all, events and the expected value of some random variables. She behaves as if she uses this information in a fashion dictated by the fat-free independence. That is, she first uses her information to form a complete preference order over a subset of acts, and then to extend it to a complete order over all acts.

In order to illustrate the last statement and the use the decision maker makes of available information, consider a gamble that guarantees the decision maker a prize whose utility is 1, if a red ball is randomly drawn form Ellsberg's urn. Denote this lottery as $(1R, 0B, 0W)$. Since the probability of red is known to be $1/3$, the expected utility of this gamble is known

[2]The reason why the property of being fat-free is too weak and why the stronger version of being strongly fat-free is necessary to exhibit the non-synergetic effects, is explained in the sequel.

to be 1/3. For a similar reason, the expected utility of a gamble denoted $(0R, 1B, 1W)$, which guarantees a prize whose utility is 1 if a either white or a black ball is drawn, is 2/3. Due to linearity, the expected value of any mixture of these two gambles is known to the decision maker. For instance, the value of $(.7R, .3B, .3W)$ is 1.3/3. However, there are many gambles – for instance $(.7R, .3B, .4W)$ – whose values are unknown to her.

In case the decision maker needs to evaluate an act whose value is not provided by the partially-specified probability, she approximates it by acts with the values of which she is acquainted. For example, as the best familiar act that approximates[3] $(.7R, .3B, .4W)$ is $(.7R, .3B, .3W)$, she evaluates $(.7R, .3B, .4W)$ as 1.3/3.

The partial information about the underlying probability is given to the decision maker through the probability of some events and the expected value of some random variables. This formulation stands in contrast with the conventional way available information about a distribution is modeled; it is usually modeled as a probability specified over a collection of events (for example, an algebra of events in de Finetti (1970) and a λ-system of events in Epstein and Zhang (2002)). To motivate this formulation, consider the following dynamic version of Ellsberg's urn. At the inception of the process there are 30 red balls and 60 white or black balls in an urn. However, the balls multiply over time at a known rate. After a while, the frequency of the red balls is no longer one third. As the process evolves, the probability of no event, but the full one, is known. This does not mean that any distribution of colors is possible. The decision maker can deduce the expected value of some random variables (for more details see Example 5 below), and thereby reduce the set of possible distributions. Based on this restricted information, the decision maker can choose between several gambles whose prizes depend on drawing random balls from the urn.

The second part of the paper extends the partially-specified probability model from one-person decision problems to strategic interactions. It introduces the concept of *partially-specified equilibrium*. In a partially-specified equilibrium, players do not have precise knowledge of the mixed strategy played by each of the other players. Rather, players know only the probability of some subsets of pure strategies, but do not know the precise sub-division of probabilities within those subsets. They might know also the expected value of some variables that depend on the actions taken by others. In other words, the mixed strategies played (which constitute probability distributions over pure strategies) are only partially specified. Moreover, different players may obtain different specifications of the mixed strategies employed by other individuals. Partially-specified equilibrium extends Nash equilibrium to cases where players best respond to partially-specified strategies. When the

[3]A precise description of how to approximate an unknown act by known ones is given in Section III.

information of all players is complete, the two notions coincide.

The model of partially-specified probability enables one to endogenize ambiguity. In partially-specified equilibrium the ambiguity players might have regarding the actual strategies played by others is not exogenous; it arises from a partial information obtained about these strategies.

In order to illustrate the idea behind the notion of partially-specified equilibrium, consider commuters who use the same system of roads. The routes taken by all drivers determine traffic conditions and thereby the utility of any individual driver using the same road system. In Nash equilibrium of this strategic situation, any driver takes the best route given traffic conditions. It implicitly assumes that each driver is fully familiar with traffic conditions. However, the information a typical commuter might hear on the radio about traffic conditions might refer only to main roads, certainly not to all of them. The implicit assumption that underlies Nash equilibrium is not fulfilled, since a commuter is unable to respond to actual traffic conditions.

A typical commuter plans her optimal route based on the partial information she possesses about the traffic situation. In partially-specified equilibrium, each commuter takes the best possible route, given the partial specifications about the actual traffic conditions that she obtains. A general ambiguity model would not specify how drivers form their beliefs about traffic conditions when they start up their engines. Here however, they get some relevant information, typically partial, form their beliefs and only then choose their route. In other words, the belief is endogenized to the model: it is directly derived from the actual strategies taken by others.

Beyond having the ability to endogenize ambiguity into strategic settings, the model of decision making under partially-specified probability, offers a few additional advantages. First, regardless of its characteristics, it always guarantees the existence of partially-specified equilibrium. The reason being that it satisfies ambiguity aversion. This, in turn, implies that the best-response correspondence is convex-valued, and therefore lends itself to a fixed-point theorem.

The second advantage is that the ambiguity present in a partially-specified equilibrium has an origin well connected to its context. Each player obtains a partial specification about the actual strategies of all other players. Based on this information, each player forms beliefs about others' strategies. Thus, the beliefs of players about others stem from the actual strategy-profile played and the partial information they obtain about it.

This leads to the third advantage. In the existing notions that extend Nash equilibrium to non-standard expected utility maximization setups, players hold some beliefs about other players' strategies. These beliefs may come either in the form of capacity[4] (see e.g., Dow and Werlang (1994) and Marinacci (2000)) or as a set of probability distributions (see e.g., Klibanoff (1996) and Lo (1999)). In these notions, there is no intrinsic connection between the actual strategies and the beliefs. This is the reason why an additional consistency condition, that relates the beliefs with the actual strategy-profile, must be imposed. For instance, in Klibanoff (1996) the set of strategies representing the belief of a player is required to contain the actual strategy-profile. In a partially-specified equilibrium, however, no additional consistency condition is required. This is so because the ambiguity stems from the actual strategies played, and therefore, the true strategy-profile is automatically contained in the belief of every player.

The last advantage, which will not be pursued in this paper, is that partially-specified probabilities can be easily adapted to incomplete information games and to cases where the actions of the players might be correlated, as in Aumann (1974).

The paper is organized as follows: Section I presents the Ellsberg paradox as a motivating example. Section II presents the model and its axioms. Section III introduces partially-specified probabilities and the integration w.r.t. them. Section IV provides the main theorem: an axiomatization of decision making with a partially-specified probability. Section V elaborates on the similarities and differences between Anscombe and Aumann (1963), Gilboa and Schmeidler (1989) and the present model. The proof of the main theorem is provided in Section VI. The notion of Partially-specified equilibrium is introduced in Section VII. Section VIII offers a discussion of the axioms. Section IX concludes with some final comments on the connection between partially-specified probabilities, Choquet integral and cooperative game theory.

I. The Ellsberg paradox - a motivating example

Suppose that an urn contains 30 red balls and 60 additional balls that are either white or black. A ball is randomly drawn from the urn and a decision maker is given a choice between the following two gambles:

Gamble **X**: Receive \$100 if a red ball is drawn.

Gamble **Y**: Receive \$100 if a white ball is drawn.

Or the decision maker is given the choice between the following two gambles:

Gamble **Z**: Receive \$100 if a red or black ball is drawn.

[4] A capacity v is a real function defined on the power set of S such that $v(\emptyset) = 0$ and $v(S) = 1$.

Gamble **T**: Receive \$100 if a white or black ball is drawn.

It is well-documented that most people strongly prefer Gamble **X** to Gamble **Y**, and Gamble **T** to Gamble **Z**. These preferences exhibit a violation of the expected utility theory.

Three states of nature are at play in this scenario: R, W and B, corresponding to one of the three colors. Denote by S the set containing these states. Each of the gambles corresponds to a real function (random variable) defined over S. For instance, Gamble **X** corresponds to the random variable X, defined as $X(R) = 100$ and $X(W) = X(B) = 0$. Let Y, Z and T be the functions that correspond to **Y**, **Z** and **T**, respectively.

Denote $\mathcal{A} = \{\emptyset, S, \{R\}, \{W, B\}\}$. There is a probability of drawing a black ball, but it is unknown to the decision maker. Only the probabilities of the four events in the sub-algebra $\mathcal{A}$ are known to the decision maker: $P(\emptyset) = 0$, $P(S) = 1$, $P(\{R\}) = \frac{1}{3}$ and $P(\{W, B\}) = \frac{2}{3}$. In other words, probability P is partially-specified.

The random variable X can be expressed as a linear combination of characteristic functions of events for which the probabilities are specified. Using only the four events in $\mathcal{A}$, X can be decomposed as $X = 100 \cdot \mathbb{I}_{\{R\}}$.[5] This decomposition is used to evaluate X at $X = 100P(R) = 100 \cdot \frac{1}{3}$.

In contrast, one cannot obtain a precise decomposition of Y. The maximal non-negative function that is below Y, and which can be written solely in terms of the events in $\mathcal{A}$, is $0 \cdot \mathbb{I}_S$. The function Y is therefore evaluated at 0. Since $100 \cdot \frac{1}{3} > 0$, **X** is preferred to **Y**.

A similar method applied to Z and T yields $Z \geq 100 \cdot \mathbb{I}_{\{R\}}$ where the right-hand side of the inequality is at its maximum value for this characteristic function. Thus, the evaluation of Z is $100 \cdot \frac{1}{3}$, while T is decomposed as $100 \cdot \mathbb{I}_{\{W,B\}}$. Therefore, the evaluation of T is $100 \cdot \frac{2}{3}$. Since $100 \cdot \frac{1}{3} < 100 \cdot \frac{2}{3}$, Gamble **T** is preferred to Gamble **Z**.

According to the intuition behind this result, the decision maker bases her evaluation of the random variables only on known quantities, namely the probability of those events for which the probability is specified. The best estimate is provided therefore by the maximal function that is not larger than the random variable and can be expressed in terms of these events. The evaluation of the random variable is based on this estimate.

[5] $\mathbb{I}_A$ is the indicator of a set A, known also as the characteristic function of A.

II. Model and axioms

A. The model

Let N be a finite set of outcomes and let $\Delta(N)$ be the set of distributions over N. S is a finite state space. Denote by L the set of all functions from S to $\Delta(N)$ and by L_c the set of all constant functions in L. An element of L is called an *act*. The constant function that attains the value y will be denoted by $\mathbf{y}$ (i.e., $\mathbf{y}(s) = y$ for every $s \in S$). L is a convex set: if $\alpha \in [0,1]$, $f, g \in L$, then $(\alpha f + (1-\alpha)g)(s) = \alpha f(s) + (1-\alpha)g(s)$.

The preference order of a decision maker is described by a binary relation $\succsim$ over L. We say that $\succsim$ is *complete* if for every f and g in L either $f \succsim g$ or $g \succsim f$. The preferences are *transitive* if for every f, g and h in L, $f \succsim g$ and $g \succsim h$ imply $f \succsim h$. The order $\succsim$ is non-trivial if there are two acts f and g such that $f \succ g$.

B. Axioms

(i) Weak Order: *The relation $\succsim$ is non-trivial, complete and transitive.*

The relation $\succsim$ defined over L induces a binary relation[6], $\succsim$, over $\Delta(N)$ as follows: $y \succsim z$ iff $\mathbf{y} \succsim \mathbf{z}$. The relation $\succsim$ induces the binary relations $\succ$ and $\sim$: $f \succ g$ iff $f \succsim g$ and not $g \succsim f$; $f \sim g$ iff $f \succsim g$ and $g \succsim f$.

Let f and g be two acts. We denote $f \geq g$ when $f(s) \succsim g(s)$ for every $s \in S$ and $f > g$ when $f \geq g$ and $f(s) \succ g(s)$ for at least one $s \in S$. For every $f \in L$, denote $W(f) = \{g;\ f \geq g\}$.

(ii) Monotonicity: *For every f and g in L, if $f \geq g$, then $f \succsim g$.*

Axioms (i) and (ii) imply that there are two constant acts $\mathbf{c}_1$ and $\mathbf{c}_2$ such that $\mathbf{c}_1 \succ \mathbf{c}_2$.

Definition 1 *(1) An act f is* fat-free *(denoted* FaF*) if $f > g$ implies $f \succ g$.*
(2) An act f is strongly fat-free *(denoted* SFaF*) if $\alpha f + (1-\alpha)\mathbf{c}_1$ is FaF for every* $0 \leq \alpha \leq 1$.

An act f is fat-free if when a lottery assigned by f to a state is replaced by a worse one, the resulting act is strictly inferior to f. In a fat-free act, there is no single lottery that can be reduced while maintaining the quality of the act. In contrast, when an act is not

[6] At the risk of creating some confusion, the same notation is used to denote the two preference orders defined over $\Delta(N)$ and L_c, though they in fact differ from one another.

fat-free, there is at least one state whose corresponding lottery can be downgraded without affecting its quality. The extra quality of this lottery is referred to as fat: it can be cut without reducing the overall value of the act.

In a particular act it might be that a lottery that corresponds to a certain state contains fat. This fat has no contribution to the value of the act. It might be, however, that when this act is mixed with another, the same fat has a significant contribution to the quality of the mixture. This might happen when the fat of two acts react synergetically with each other. While each, separately, does not contribute to the value of its corresponding act, when combined together they add an extra value.

The role of the specific independence axiom adopted here is to avoid also this synergetic effect. It turns out that considering fat-free acts is not sufficient for this purpose. The independence axiom should involve acts that are not only fat free, but also do not exhibit any synergetic effect when combined with any other act. These are the strong fat-free acts.

Testing whether an act is fat-free or not is done by downgrading the lotteries involved, and checking whether the quality of the act is harmed or not. When a lottery associated with a particular state is already the worst possible, there is no way to downgrade it, meaning that there is nothing to check. However, the fact that a lottery cannot be downgraded does not mean that it is free of synergetic effects. If it could be – theoretically – reduced further, it would damage the value of the act. In this case I figuratively refer to lottery as containing hidden fat. The hidden fat contained in a lottery can react synergetically with the fat contained in another act.

Hidden fat can be revealed by (a) uniformly change all the lotteries associated with an act in order to make them all strictly preferred to the worst one; and (b) checking whether after this change the act is still fat-free. If after uniformly changing an act it becomes non fat-free, then it contains hidden fat. An act with no hidden fat is called strongly fat-free. More formally, the test of whether f is strongly fat-free or not is carried out by mixing it with a constant act $\mathbf{c}_1$: $\alpha f + (1-\alpha)\mathbf{c}_1$ in order to make all the lotteries involved strictly preferred to the worst one. If $\alpha f + (1-\alpha)\mathbf{c}_1$ is fat-free for every $\mathbf{c}_1$, the act is strongly fat-free. The fact that $\alpha f + (1-\alpha)\mathbf{c}_1$ is fat-free implies that fat was not found, not because one of the lotteries hit rock bottom, but rather because there is no hidden fat.

According to the definition of fat-free acts, if a lottery related to a specific state can be downgraded without changing the value of the act, then the act bears fat. In this definition there is no reference to the lotteries related to other states: whether or not fat exists depends only on the lottery related to a specific state and to other states. In contrast, the existence of hidden fat in a specific state depends on other states as well. When a lottery of a specific state hits rock bottom and cannot be downgraded further, this does not mean that there is no hidden fat. There might be hidden fat, but this fact cannot be verified. However, if all lotteries are changed (by mixing with constant $\mathbf{c}_1$) and the act becomes non fat-free, this is a testimony to the existence of hidden fat. This hidden fat can be revealed only when all the lotteries involved become strictly better than the worst lottery, and therefore there is a possibility to downgrade each. In other words, the existence of hidden fat in a particular state involves changing also the lotteries of other states.

Remark 1 Lemma 4 below shows that the definition of SFaF does not depend on the choice of $\mathbf{c}_1$, as long as there exists $\mathbf{c}_2$ that satisfies $\mathbf{c}_1 \succ \mathbf{c}_2$ (for further discussion, see subsection VIII C.

The notions presented in Definition 1 are clarified by the next two examples.

Example 1 Consider Ellsberg's urn described in Section I. Suppose that N is the set of integers between 0 and 100. Act X, which takes the value of 100 on R and 0 on the rest, is FaF since a reduction in any of the prizes results in a worse act. For instance, X', which takes the value of 99 on R and 0 on the rest, is worse than X. On the other hand, Y is not FaF since Y', which coincides with Y on R and B and is equal to 99 on W, is equivalent to Y.

Example 2 An urn contains 100 balls of four colors: white, black, red and green. It is known that there are 90 white or black balls and that there are 90 white or red balls. Thus, $S = \{W, B, R, G\}$, $P(W, B) = P(W, R) = .9$. Consider the act $\mathbb{I}_{\{W,B\}}$ which takes the value of 1 on $\{W, B\}$ and 0 otherwise. The expected value of this act is .9. Moreover, the expectation of any function of type $Z = \alpha\mathbb{I}_{\{W,B\}} + (1 - \alpha)\mathbb{I}_S$ is $.9\alpha + (1 - \alpha)$ and the expectation of any act smaller than Z is strictly smaller than the expectation of Z. Thus, $\mathbb{I}_{\{W,B\}}$ is SFaF since Z can be expressed by acts whose expectation is known.

Now consider act $\mathbb{I}_W$, which takes the value of 1 on W and 0 otherwise. The information available tells us nothing about the probability of W; it can be anywhere between .8 and .9. A decision maker who dislikes ambiguity would evaluate the expectation of $\mathbb{I}_W$ at .8. Adding $\mathbb{I}_S$ to $\mathbb{I}_W$ would result in an act, say Y, whose expectation is 1.8.

Denote $X = \mathbb{I}_{\{W,B\}} + \mathbb{I}_{\{W,R\}}$. Note that the expectation of X is also 1.8. However, X coincides with Y in all states except G. The act X takes the value 0 on G while Y takes the value 1. Thus, reducing the value that Y takes on G from 1 to 0 does not reduce the expectation of Y. Therefore, $\mathbb{I}_W$ is not SFaF. This is because there is no way to obtain a precise evaluation of the probability of W. Instead, an approximate evaluation is obtained by taking the lowest estimation that is still consistent with the information provided. ■

(iii) Strongly Fat-Free Independence: *Let f and g be acts. Then, $f \succ g$ if and only if $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$ for every h which is $SFaF$ and $\alpha \in (0,1)$.*

The idea behind the independence axiom is that mixing f and g with another act does not interfere with the preference order. That is, $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$ as long as $f \succ g$. The classical independence axiom of Anscombe and Aumann (1963) requires that this property holds for every act h. Gilboa and Schmeidler (1989), on the other hand, restrict h to be the simplest possible, meaning a constant. Here, we adopt an axiom that lies in between these two versions. It requires that no strongly fat-free interferes with the preference order. This axiom suggests that if mixing with an act h flips the preference order (i.e., $f \succ g$ while $\alpha f + (1-\alpha)h \preceq \alpha g + (1-\alpha)h$), it is due to the fat contained in h. This phenomenon happens when the fat in h and g react synergetically with each other. The preference order is thus reversed when mixing g with h creates a synergetic effect that is stronger than that between h and f.

The following is the standard continuity axiom.

(iv) Continuity: *For every f, g and h in L: (a) if $f \succ g$ and $g \succsim h$, then there exists α in $(0,1)$, such that $\alpha f + (1-\alpha)h \succ g$; and (b) if $f \succsim g$ and $g \succ h$, then there exists β in $(0,1)$, such that $g \succ \beta f + (1-\beta)h$.*

For the sake of simplicity, the following conciseness axiom is added. It states that every constant act is FaF. If a constant act is not FaF, then one state is known to have probability zero, in which case this state can be omitted from S.

(v) Conciseness: Any constant act is FaF.

III. Partially-specified probabilities

The following model of decision making with partially-specified probabilities illustrates the case where the decision maker obtains some data regarding the *true* expectation of some of the random variables. A *partially-specified probability* over S is a pair $(P, \mathcal{Y})$, where $\mathcal{Y}$ is a set of real functions over S, $\mathcal{Y}$ contains $\mathbb{I}_S$ and P is a probability over S.

The decision maker obtains a partial information about the probability P: he knows the value $E_P(Y)$, which is the expectation of Y w.r.t. P, for every $Y \in \mathcal{Y}$. The larger $\mathcal{Y}$ is, the richer the information. If, for instance, $\mathcal{Y}$ contains all the real functions over S, then P is fully-specified, while if it contains only $\mathbb{I}_S$ no information about P is provided to the decision maker.

Since the decision maker knows $E_P(Y)$ for every $Y \in \mathcal{Y}$, she can calculate the expectation of any function of the form $\sum_{Y\in\mathcal{Y}} \lambda_Y Y$. We use this fact in order to define the integral of any non-negative function w.r.t. a partially-specified probability.

Let ψ be a non-negative function defined over S and let $(P, \mathcal{Y})$ be a partially-specified probability. Denote:

$$(1) \quad \int \psi dP_{\mathcal{Y}} = \max\Big\{ \sum_{Y\in\mathcal{Y}} \lambda_Y E_P(Y); \ \sum_{Y\in\mathcal{Y}} \lambda_Y Y \leq \psi \text{ and } \lambda_Y \in \mathbb{R} \text{ for every } Y \in \mathcal{Y}\Big\},$$

The integral of ψ, inspired by the concave integral for capacities (Lehrer, 2009), is the maximal $\sum_{Y\in\mathcal{Y}} \lambda_Y E_P(Y)$ among all those functions of the form $\sum_{Y\in\mathcal{Y}} \lambda_Y Y$ that are below ψ.

Although the set $\mathcal{Y}$ need not be closed, the maximum (as opposed to the supremum) on the right-hand side of eq. (1) is obtained. Indeed, the following lemma states that without loss of generality $\mathcal{Y}$ can be assumed to be finite. The employment of the word "maximum" is therefore justified.

Lemma 1 *Let $(P, \mathcal{Y})$ be a partially-specified probability. Then, there is a finite subset of $\mathcal{Y}$, say $\mathcal{Y}'$, such that*

$$\int \psi dP_{\mathcal{Y}} = \int \psi dP_{\mathcal{Y}'}.$$

The following two examples illustrate how such integration is carried along.

Example 3 Let $S = \{s_1, s_2, s_3, s_4\}$ and suppose that $\mathcal{Y} = \{\mathbb{I}_S, \mathbb{I}_{\{s_1,s_2\}}, \mathbb{I}_{\{s_3,s_4\}}\}$. Furthermore, assume that $P(s_1, s_2) = \frac{1}{3}$ and $P(s_3, s_4) = \frac{2}{3}$. Consider $\psi = (1, 2, 3, 4)$, i.e. $\psi(s_i) = i,\ i = 1, 2, 3, 4$. The random variable ψ is greater than or equal to $\mathbb{I}_{\{s_1,s_2\}} + 3\mathbb{I}_{\{s_3,s_4\}}$. This linear combination maximizes the right-hand side of eq. (1) and therefore, $\int \psi dP_{\mathcal{A}} = \frac{1}{3} + 3\frac{2}{3} = \frac{7}{3}$. ■

Example 4 Consider $S = \{s_1, s_2, s_3, s_4\}$ and suppose that the real probability over S is P. However, the decision maker is informed only of the probability of $A = \{s_1, s_2\}$ and $B = \{s_2, s_3\}$ (and, as usual, of S). Note that A, B and S do not form an algebra. In this case, $\mathcal{Y} = \{\mathbb{I}_A, \mathbb{I}_B, \mathbb{I}_S\}$. Let $\psi = \mathbb{I}_{S\setminus A}$. Since, $\mathbb{I}_{S\setminus A} = \mathbb{I}_S - \mathbb{I}_A$ we obtain $\int \psi dP_{\mathcal{Y}} = P(S) - P(A)$.

Now let ψ be a function over S defined as follows: $\psi(s_1) = 1 - \alpha$, $\psi(s_2) = \alpha$, $\psi(s_3) = 1 - \alpha$ and $\psi(s_4) = 2 - 3\alpha$, where $\alpha \leq \frac{1}{2}$ (and thus, $2 - 3\alpha \geq 1 - \alpha \geq \alpha$). Notice that $(1-\alpha)\mathbb{I}_{S\setminus A} + (1-\alpha)\mathbb{I}_{S\setminus B} - \alpha\mathbb{I}_S \leq \psi$. Therefore, $\int \psi dP_{\mathcal{Y}} \geq (1-\alpha)P(S\setminus A) + (1-\alpha)P(S\setminus B) - \alpha$.

Also, $\alpha\mathbb{I}_A + (1-\alpha)\mathbb{I}_{S\setminus A} \leq \psi$ and therefore, $\int \psi dP_{\mathcal{Y}} \geq \alpha P(A) + (1-\alpha)P(S \setminus A)$. It turns out that $\int \psi dP_{\mathcal{Y}} = \max((1-\alpha)P(S\setminus A) + (1-\alpha)P(S\setminus B) - \alpha, \alpha P(A) + (1-\alpha)P(S\setminus A))$. If

$\alpha = 0$, for instance, then $\int \psi dP_{\mathcal{Y}} = \max(P(S\backslash A)+P(S\backslash B), P(S\backslash A)) = P(S\backslash A)+P(S\backslash B)$. This example is revisited in the context of strategic interactions (see Example 7 below). ∎

In the most intuitive examples, such as the Ellsberg urn and Example 3 above, the governing probability is specified over an algebra of events. The following example describes a situation whereby the provided information contains the true expected value of some random variables, but not the probability of any event except for the entire state space.

Example 5 (Dynamic Petri dish) Suppose that, as in an Ellsberg urn, on day 1 a Petri dish contains 90 organisms of which 30 are red and the rest are white or black. However, each white organism splits into two once per day. On day 2 the decision maker is called upon to choose between a number of gambles.

The information available to the decision maker on day 2 does not contain the probability of any event, and furthermore the probability of R is no longer $\frac{1}{3}$. However, the decision maker can deduce the expectation of certain random variables from the information available. To illustrate the point, suppose that the numbers of red, white and black organisms on day 2 are denoted by n_r, n_w and n_b, respectively. On day 1, there were $\frac{n_w}{2}$ white organisms. It is known that $\frac{n_r}{n_r+\frac{n_w}{2}+n_b} = \frac{1}{3}$. Since $\frac{n_w/6}{n_w/2} = \frac{1}{3}$ one obtains $\frac{n_r+n_w/6}{n_r+\frac{n_w}{2}+n_b+n_w/2} = \frac{n_r+\frac{1}{6}n_w}{n_r+n_w+n_b} = \frac{1}{3}$. This is the expectation (on day 2) of the random variable that receives the value 1 on Red, $\frac{1}{6}$ on White and 0 on Black. In other words, according to the information available, the expectation of this random variable is $\frac{1}{3}$.

The decision maker knows that a regular probability, i.e. the actual distribution of organisms, underlies this decision problem. However, she obtains only partial information regarding it. Beyond the obvious information concerning the probability of the whole space and the empty set, on day 2 the decision maker is informed only of the expectation of the random variable $(1, \frac{1}{6}, 0)$, from which she can also deduce the expectation of any random variable of the form $c(1, \frac{1}{6}, 0)$, $c > 0$. No further information is available. ∎

An important feature of this integral is concavity, which means that the integral, as a function defined on the space of functions, is concave.

Lemma 2 *Let $(P, \mathcal{Y})$ be a partially-specified probability. Then, for any two functions ψ and ϕ,*

$$\int \alpha\psi + (1-\alpha)\phi dP_{\mathcal{Y}} \geq \alpha \int \psi dP_{\mathcal{Y}} + (1-\alpha) \int \phi dP_{\mathcal{Y}}.$$

The following lemma links between the partially-specified probability model and the multiple-prior model (Gilboa and Schmeidler, 1989). It states that for every partially-specified probability $(P, \mathcal{Y})$ one can find a finite number of probability distributions, such that the integral with respect to $(P, \mathcal{Y})$ is equal to the minimum of the respective (additive) integrals.

Lemma 3 *Let $(P, \mathcal{Y})$ be a partially-specified probability. Then, there is a finite subset $\mathcal{Q}$ of probability distributions over S such that*

$$\int \psi dP_{\mathcal{Y}} = \min_{Q \in \mathcal{Q}} \int \psi dQ.$$

We now turn to the characterization of the integral w.r.t. a partially-specified probability.

Definition 2 *Let I be a real function over $[0,1]^S$ and let $X, Y \in [0,1]^S$.*
(i) We say that $X \geq Y$ if $X(s) \geq Y(s)$ for every $s \in S$ and $X > Y$ if $X \geq Y$ and $X(s) > Y(s)$ for at least one $s \in S$.
(ii) A function X over S is fat-free *(FaF) w.r.t. I if $X > Y$ implies $I(X) > I(Y)$.*
(iii) A function X over S is strongly fat-free *(SFaF) w.r.t. I, if $\alpha X + (1-\alpha)\boldsymbol{c}$ is FaF w.r.t. I for every $0 < \alpha \leq 1$ and every constant function $\boldsymbol{c}$.*

Given a function X over S, $[X]$ denotes a function, which is FaF w.r.t. I, $X \geq [X]$ and $I(X) = I([X])$.

Proposition 1 *Let I be a real function over $[0,1]^S$. There is a partially-specified probability, $(P, \mathcal{Y})$, such that $I(X) = \int X dP_{\mathcal{Y}}$ iff*
(1) *I is monotonic w.r.t. to $\geq$;*
(2) *For every X, there is $0 \leq \alpha < 1$ such that $[\alpha X + (1-\alpha)\mathbb{I}_S]$ is SFaF w.r.t. I;*
(3) *If X is SFaF, Y is a function and $\alpha \in (0,1)$, then $I(\alpha X) + I((1-\alpha)Y) = I(\alpha X + (1-\alpha)Y)$;*
(4) *For every X and for every positive c, if $cX \in [0,1]^S$, then $I(cX) = cI(X)$;*
(5) *$c\mathbb{I}_S$ is SFaF w.r.t. I for every $c \in [0,1]$; and*
(6) *$I(\mathbb{I}_S) = 1$;*

IV. Decision making with a partially-specified probability

Theorem 1 *Let $\succsim$ be a binary relation over L. This satisfies (i)-(iv) and (v) if and only if there is a partially-specified probability $(P, \mathcal{Y})$ with finite $\mathcal{Y}$ and an affine function u on $\Delta(N)$, such that for every $f, g \in L$,*

$$f \succsim g \qquad \text{iff} \qquad \int u(f(s))dP_{\mathcal{Y}} \geq \int u(g(s))dP_{\mathcal{Y}}, \tag{2}$$

and, moreover, if $\boldsymbol{c}$ is a constant act and $f \leq \boldsymbol{c}$ satisfies $\int u(f(s))dP_{\mathcal{Y}} \geq \int u(\boldsymbol{c})dP_{\mathcal{Y}}$, then $f = \boldsymbol{c}$.[7]

In order to evaluate an act, a decision maker uses the information captured in $(P, \mathcal{Y})$. Anything that can be deduced from this information is employed to the full extent. For instance, if the probability of an event is provided, then the probability of its complement can be deduced. Furthermore, a decision maker can deduce the expectation of any linear combination of random variables whose expectations are known.

The model of decision making described in Theorem 1 is consistent with the multiple-prior model (Gilboa and Schmeidler, 1989) rather than with the model of Choquet expected-utility maximization (Schmeidler, 1989). However, two features of the current model make it more structured than the multiple-prior model. First, due to Lemma 1, the decision maker uses only a finite number of priors to determine his preference order over acts. Second, there is one probability distribution (referred to as "real") and a finite set of random variables on which all the priors and the real distribution agree. Moreover, any other distribution that satisfies this last condition can be obtained as a linear combination of the priors and any distribution which is a linear combination of the priors satisfies this condition. It is important to emphasize that a multiple-prior model with a polytope of priors is typically not a partially-specified probability model. In order to be a partially-specified probability the set of priors needs to be a polytope of a special kind: an intersection of an affine space (determined by a finite set of random variables and their expected values) and the set of all priors.

The Choquet expected-utility maximization model and the multiple-prior model are belief-based. In other words, they assume that there is a (non-additive) belief or a set of priors (beliefs) that the decision maker uses to construct her preferences. In contrast, the

[7] Throughout this paper equality between acts is used in a broad sense. Thus, if we let f and g be two acts, we say that $f = g$ if for every $s \in S$, $f(s) \sim g(s)$.

partially-specified probability model is information-based. In other words, it assumes that the decision maker obtains partial information regarding the real distribution and based on this information constructs her preferences. While the belief-based models hinge on a fixed belief or on a fixed set of priors, the information-based model allows the belief to change with the underlying (real) distribution.

This difference between the models is crucial when the underlying (i.e., the real) distribution can vary. It is particularly important in a strategic interaction when the states of nature are nothing but other players' actions and the source of uncertainty is the scheme according to which other players are mixing or are randomly matched with each other. The partially-specified probability model enables one to examine a strategic interaction with a given information structure (which specifies what each player knows about other players' strategies) and to analyze the possible equilibria in such a situation. Different equilibria determine different beliefs that players hold about the interaction. In other words, whereas the information available is fixed, beliefs can change. This issue is also discussed in the context of partially-specified equilibrium (see Section VII).

Notice that Gilboa and Schmeidler (1989)'s uncertainty aversion axiom does not play a role in Theorem 1. This axiom is not assumed but rather implied, as stated in the following corollary.

(vi) Uncertainty Aversion: *For every f, g and h in L, if $f \succsim h$ and $g \succsim h$, then for every α in $(0,1)$, $\alpha f + (1-\alpha)g \succsim h$.*

Corollary 1 *Let $\succsim$ be a binary relation over L. Then, axioms (i)-(iv) and (v) imply (vi).*

This corollary is a consequence of Theorem 1 and Lemma 2, which states that the integral w.r.t. a partially-specified probability is a concave operator defined on the set of functions. Note that the integral w.r.t. a partially-specified probability is not the only concave operator involved in decision theory. Choquet integral w.r.t. a convex capacity or w.r.t. a belief function (see, Dempster (1967) and Shafer (1976)) are also concave operators (see, Azrieli and Lehrer (2007a), Proposition 2). However, the integral w.r.t. a partially-specified probability typically cannot represent or be represented neither by Choquet integral w.r.t. a convex capacity, nor by Choquet integral w.r.t. a belief function.

V. Expected utility, partially-specified probability and maxmin expected utility

This section is devoted to a comparison between expected utility maximization, Partially-specified probability and maxmin expected utility maximization.

A decision maker who maximizes expected utility does it with respect to one probability distribution. On the other hand, a decision maker who follows the maxmin expected utility model of Gilboa and Schmeidler (1989) uses a set of probability distributions. Partially-specified probabilities generate a special kind of sets of distributions. These are sets consisting of all distributions that agree about the expected values of a list of random variables.

Example 2 continued Consider again Example 2. There, the state-space is $S = \{W, B, R, G\}$ and the information available is $P(W, B) = P(W, R) = .9$. In terms of distributions, it means that the true distribution over S assigns probability .9 to the events $\{W, B\}$ and $\{W, R\}$. In other words, the true distribution is $(P(W), P(B), P(R), P(G))$, which satisfies the two regular conditions that define a distribution (i.e., $P(W) + P(B) + P(R) + P(G) = 1$ and $P(i) \geq 0, i \in S$) and two additional ones: $P(W) + P(B) = .9$ and $P(W) + P(R) = .9$. For instance, distributions $(.9, 0, 0, .1)$ and $(.8, .1, .1, 0)$ satisfy these conditions, while $(.8, .1, 0, .1)$ does not.

The set of distributions consistent with this information consists of all the distributions that cannot be refuted by it. It consists of the distributions that satisfy all the equalities and inequalities specified above. This set is one-dimensional because there are four variables that are subject to three equations. This set is explicitly described as $\{(.9 - \beta, \beta, \beta, .1 - \beta);\ 0 \leq \beta \leq .1\}$. Once again, this distributions' set was derived from a list of random variables and their expectations, and is therefore equivalent to a partially-specified probability.

Would the information available be richer and include, for instance, $P(G) = 0$, the set of consistent distribution would reduce to a singleton whose member is $(.8, .1, .1, 0)$. In such a case, taking a decision with respect to the partially-specified probability would be equivalent to expected utility maximization.

The question arises as to whether any set of distributions can be described by a list of linear equations. If the answer to this question would be on the affirmative, the decision model described here would be equivalent to the maxmin expected utility model of Gilboa and Schmeidler (1989). The answer, however, is negative.

Consider the set of distributions $(P(W), P(B), P(R), P(G))$ that satisfy $.6 \leq P(W) \leq 1$. This set is not the solutions set of linear equations and cannot be derived from a partially-specified probability. We will prove it by showing that the preference order induced by maxmin expected utility maximization based on this set of priors does not satisfy axiom (iii).

Consider two actions, f and g, that correspond to the utility vectors $f' = (.5, 0, 0, 0)$ and $g' = (0, 1, 1, 1)$, resp. As far as f is concerned, the worst prior is $P = (.6, 0, 0, 0)$, whereas when g is concerned, the worst is $Q = (1, 0, 0, 0)$. Since the expected value of f' w.r.t. P is .3, and that of g' w.r.t. Q is 0, f is strictly preferred to g. Now consider the act h that corresponds to the utility vector $h' = (1, 0, 0, 0)$. This act is strongly fat-free. The mixtures $.5f + .5h$ and $.5g + .5h$ correspond to $(.75, 0, 0, 0)$ and $(.5, .5, .5, .5)$. The first, $.5f + .5h$, is equivalent (according to the maxmin expected utility maximization model) to $(.45, .45, .45, .45)$, and therefore is strictly preferred by $.5g + .5h$. This means that mixing with a strong fat-free act flips the preference order over f and g, and therefore violates axiom (iii). ■

To make the connection between general ambiguity and the current model more intuitive, let $(P, \mathcal{Y})$ be a partially-specified probability. Denote by span$(\mathcal{Y})$ the linear space spanned by $\mathcal{Y}$. It consists of all linear combinations of variables in $\mathcal{Y}$. Since the value of $E_P(Y)$ is known for every $Y \in \mathcal{Y}$, so is $E_P(Y')$ for every $Y' \in$ span$(\mathcal{Y})$. That is, the decision maker knows not only the expected value of any function in $Y \in \mathcal{Y}$, but also of those in span$(\mathcal{Y})$.

The knowledge of the decision maker does not restrict the compatible-with-information distributions to only one. While the information about the variables in span$(\mathcal{Y})$ is provided through their expectation w.r.t. the distribution P, the same information could as well be given w.r.t another probability distribution. For instance, if P' agree with P, on every random variable in span$(\mathcal{Y})$,[8] they agree, in particular, on every variable in $\mathcal{Y}$. In other words, $(P, \mathcal{Y}) = (P', \mathcal{Y})$.

Stated differently, the source of the ambiguity is that when obtaining the partial specification about P through $(P, \mathcal{Y})$, the decision maker cannot infer whether the origin of this information is P or P'. All the probability distributions that agree with P on the variables in $\mathcal{Y}$ form the set of possible (i.e., compatible-with-information) distributions – the set of priors.

[8] In the sense that $E_P(Y')$ =E$_{P'}(Y')$.

VI. Proof of Theorem 1

A. Some implications of the axioms

The next lemma relates to the issues discussed in Remark 1. Axioms (i), (iii), (iv) and (vi) ensure the existence of a vN-M representation of $\succsim$ on L_c. Denote by $\mathbf{m}$ and $\mathbf{M}$ the minimum and the maximum in L_c, respectively. The definition of SFaF uses the constant act $\mathbf{c}_1$. The following lemma states that the definition is independent of the choice of $\mathbf{c}_1$ and that $\mathbf{M}$ can be used as well. when mixing an act f with $\mathbf{M}$ all lotteries are uniformly improved and become strictly preferred to $\mathbf{m}$, making downgrading of all lotteries possible.

Lemma 4 *Axioms (i)-(v) imply that f is SFaF if and only if for every α, $\alpha f + (1-\alpha)\boldsymbol{M}$ is SFaF.*

We formulate in addition several preliminary results, implied by the axioms, which will be helpful in proving the main result – Theorem 1. These results are proved in the appendix. The following lemma states that every act has an equivalent FaF act.

Notation 1 *Let f be an act. $[f]$ denotes an FaF act that satisfies $f \geq [f]$ and $f \sim [f]$.*

Lemma 5 *Axioms (i)-(v) imply that for every act f, there exists $[f]$.*

The next lemma states that the set of SFaF acts is convex.

Lemma 6 *Axioms (i)-(v) imply that if f and g are SFaF, then for every $0 \leq \alpha \leq 1$, $\alpha f + (1-\alpha)g$ is also SFaF. In other words, the set of SFaF acts is convex.*

Every act has an equivalent FaF act, though not an SFaF act. However, as the next lemma states, every act has some mixture with the maximally constant act that is SFaF.

Lemma 7 *Axioms (i)-(v) imply that for every act f, there is $0 < \alpha \leq 1$ such that $[\alpha f + (1-\alpha)\boldsymbol{M}]$ is SFaF.*

B. The proof of Theorem 1

The proof of the 'only if' direction is omitted. In order to prove the 'if' direction, assume that Axioms (i)-(v) are satisfied. Axioms (i), (iii), (iv) and (v) guarantee that there is a vN-M representation over L_c. Moreover, U can be normalized to take the values of 0 and 1 on $\mathbf{m}$ and $\mathbf{M}$, respectively. Axioms (ii) and (iv) ensure that every act f has an equivalent constant act, say $\mathbf{k(f)}$. Define $U(f) = U(\mathbf{k(f)})$. The function U represents $\succsim$. That is, for every two acts f, g, $U(f) \geq U(g)$ if and only if $f \succsim g$.

Denote by L_{SFaF} the set of the SFaF acts. By Lemma 6, U is affine over L_{SFaF}. Define I on $[0,1]^S$ as follows: Define $u(y) = U(\mathbf{y})$ for every $y \in \Delta(N)$. For every $X \in [0,1]^S$ and every act f, there is an act f_X such that for every $s \in S$, $X(s) = u(f_X(s))$. Note that $X_{f_X} = X$. Moreover, if X is SFaF, so is f_X. Now, define $I(X) = U(f_X)$. Thus, $I(X_f) = U(f)$ for every $f \in L$.

We show that I satisfies properties (1)-(6) of Proposition 1. (1) is ensured by (ii); (2) is guaranteed by Lemma 7; (4) is due to $I(0\mathbb{I}_S) = 0$, $0\mathbb{I}_S$ being SFaF and $\succsim$ having a vN-M representation over the set of acts $\alpha\mathbf{m} + (1-\alpha)f_X$ for every function X; (3) holds due to (4) and (iii); (5) is satisfied as a result of (vi) and due to (4); and finally, (6) is satisfied since $f_{\mathbb{I}_S} = \mathbf{M}$. Thus, $I(\mathbb{I}_S) = U(f_X) = U(\mathbf{M}) = 1$.

Proposition 1 ensures that there is a partially-specified probability $(P, \mathcal{Y})$ such that $I(X) = \int X dP_{\mathcal{Y}}$ for every $X \in [0,1]^S$. Thus, $U(f) = I(X_f) = \int X_f dP_{\mathcal{Y}} = \int u \circ f dP_{\mathcal{Y}}$. By Lemma 1, $\mathcal{Y}$ can be chosen to be finite, which is the desired result.

Moreover, suppose that $f \leq \mathbf{c}$ and $f \sim \mathbf{c}$. This implies that $X_f \leq c\mathbb{I}_S$ and that $I(X_f) = \int X_f dP_{\mathcal{Y}} = c$. However, $c\mathbb{I}_S$ is SFaF and thus, $X_f = c\mathbb{I}_S$, which implies that $f \sim \mathbf{c}$, as required. ∎

VII. Equilibrium with partially-specified probabilities

As mentioned in the Introduction, commuters plan their optimal routes in order to minimize travel time, even though they have only partial information regarding traffic conditions. The distribution of vehicles on the roads, to which an individual commuter responds, is not determined by nature, but rather by the accumulation of the choices made by many individual decision makers. However, in contrast to the traditional assumption that underlies Nash equilibrium, according to which players optimally respond to the actions taken by all other players, in this case players respond to only partial information regarding the actions of the other players. This section introduces notions of equilibria according to which players obtain only a partial specification of other players' actions.

Let $G = (M, \{B_i\}_{i \in M}, \{u_i\}_{i \in M})$ be a game in which M is the finite set of players, B_i is player i's (finite) set of pure strategies and $u_i : B \to \mathbb{R}$ is player i's payoff function, where $B = \times_{i \in M} B_i$. Denote by $\Delta(B_i)$ the set of player i's mixed strategies and for every i, denote $B_{-i} = \times_{j \neq i} B_j$. B_{-i} is the set of action profiles of all players excluding i. For every $b_{-i} \in B_{-i}$ and player i's mixed strategy $p_i \in \Delta(B_i)$, define $u_i(p_i, b_{-i}) = \sum_{b_i \in B_i} p_i(b_i) u_i(b_i, b_{-i})$. This is the linear extension of u_i to $\Delta(B_i) \times B_{-i}$.

Suppose that the only information player i has about the players $M \setminus \{i\}$ consists of the probabilities of some, though not all, subsets of B_i or of the expectation of some, though not all, of the random variables defined on B_i. Denote by $\mathcal{Y}_i^j$ the set of random variables defined over B_j whose expectations are specified to player i. More formally, suppose that player j $(j \neq i)$ plays the mixed strategy p_j. Player i does not know p_j in its entirety, but rather, knows only the expectation of each $Y \in \mathcal{Y}_i^j$. In equilibrium, player i's strategy, p_i, must be a best response to all other players' strategies, as perceived by player i.

In addition to the information a player obtains regarding other players' strategies, in partially-specified equilibrium a player also knows that all other players play independently of one another. To illustrate this point, consider the following example:

Example 6 There are three players. Players 1 and 3 have two actions each while player 2 has three. Player 1 chooses a row, player 2 a column and player 3 a matrix. The numbers indicated in the following matrices are the payoffs of player 3:

	L	M	R
T	1	0	-1
B	2	2	3

L	M	R
1	4	5
2	0	-1

Suppose that player 3 is informed that player 1 plays each row with probability $\frac{1}{2}$ but has no information regarding player 2's strategy. Based on this information and on the fact that players 1 and 2 are not coordinated, player 3 must play the left matrix with probability p that maximizes $\frac{1}{2}\min[3, 2p + 4(1-p), 2p + 4(1-p)]$, in order to maximize the worst-case payoff. The probability that does this is $\frac{1}{2}$. When the left matrix is played with probability $\frac{1}{2}$, player 3 guarantees the payoff 1.5. However, if player 3 does not take into account that players 1 and 2 are not coordinated, he must consider all distributions - not just the independent ones - that are consistent with his information. In other words, he must also consider coordinated strategies according to which player 2 plays one mixed action when player 1 plays T and another mixed action when player 1 plays B. In this case, he must play the left matrix with probability p that maximizes:

$$(3) \qquad \frac{1}{2}\Big(\min[1, 4(1-p), -p+5(1-p)] + \min[2, 2p, 3p-(1-p)]\Big).$$

By playing the left matrix with probability $p = \frac{2}{3}$, player 3 guarantees the payoff $\frac{7}{6}$. From player 3's point of view, the worst case is when player 1 plays T and B with equal probability and player 2 plays either L or R whenever player 1 plays T and M whenever player 1 plays B. This coordinated joint strategy of player 1 and 2 is consistent with player 3's information. Note that in order to bring player 3's payoff down to $\frac{7}{6}$ players 1 and 2 must play in a coordinated manner. The best they can do by playing independently of one another is to reduce player 3's payoff to 1.5. ■

Let $(p_i)_{i\in M}$, where $p_i \in \Delta(B_i)$ for every $i \in M$, be a profile of mixed strategies. Denote $p_{-i} = (p_i)_{j\neq i}$. Let $C(p_j, \mathcal{Y}_i^j)$ be the set of all player j's mixed strategies that are consistent with p_j and $\mathcal{Y}_i^j$. Formally, $C(p_j, \mathcal{Y}_i^j) = \{q_j \in \Delta(B_j);\ E_{q_j}(Y) = E_{p_j}(Y) \text{ for every } Y \in \mathcal{Y}_i^j\}$. The strategy p_i is a *best response* to p_{-i}, given $\mathcal{Y}_i^j$ $(j \neq i)$, if p_i maximizes:

$$(4) \qquad \min u_j(p_i, q_{-i}),$$

where the minimum is taken over all $q_j \in C(p_j, \mathcal{Y}_i^j),\ j \neq i$.

Definition 3 *Let $\mathcal{Y}_i^j$ be the set of random variables defined over B_j whose expectations are specified for player i. A profile $\{p_i\}_{i\in M} \in \times_{i\in M}\Delta(B_i)$ of mixed strategies is a* partially-specified equilibrium *w.r.t. $\mathcal{Y}_i^j$ $(j \neq i)$, if for every player i the mixed strategy p_i is a best response to p_{-i}, given $\mathcal{Y}_i^j$ $(j \neq i)$.*

In partially-specified equilibrium, as in Nash equilibrium, it is implicitly assumed that the players play independently of one another. This is the reason that each player maximizes her payoff against the worst possible *independent* distribution over other players' strategies that is consistent with her information. In order to evaluate the payoff associated with any possible strategy, the players do not resort to the approximating method (as in eq. (1)). Furthermore, they do not evaluate the payoffs by approximating the real payoffs using the random variables whose expectation is known, as done in Section III above. Rather, the players play mixed strategies that guarantee the highest possible payoffs against any independent distributions that are consistent with their information. The reason for adopting this particular definition is that evaluating the approximating method corresponds to guaranteeing a payoff against any distributions that is consistent with the information available, rather than only against the independent distributions.

To illustrate this point, consider Example 6 again. If player 3 plays L with probability p, the expected payoff-matrix is:

	L	M	R
T	1	4(1-p)	5-6p
B	2	2p	4p-1

From player 3's perspective, the entries of the matrix form a probability space. She is informed of the expectation of two random variables. The expectation of the random variable that is identically 1 on the top row and identically 0 on the bottom row is $\frac{1}{2}$. Likewise, the expectation of the random variable that is identically 1 on the bottom row and identically 0 on the top row is $\frac{1}{2}$. Using these random variables to evaluate the expected payoff matrix by the approximation method precisely entails finding p that maximizes (3). This corresponds to minimizing over all distributions, including the coordinated ones, that are consistent with the information, rather than just the independent distributions.

The notion of partially-specified equilibrium is information-based. The information structure, i.e. the type of information a player receives about others' strategies is exogenous. The strategies actually played by the other players and the information structure induce the set of priors that a player holds. No player arrives to the game with prior beliefs about other players' strategies. This set is endogenously determined.

When the set $\mathcal{Y}_i^j$ contains all the variables of the form $\mathbb{I}_{b_j}$ for each player i and for every $b_j \in B_j$ $(j \neq i)$, the partially-specified equilibrium w.r.t. $\mathcal{Y}_i^j$ coincides with Nash equilibrium. However, unlike Nash equilibrium, in a typical partially-specified equilibrium the pure strategies that belong to the support of player i's strategy are not necessarily best responses to p_{-i}. To illustrate this point consider the following example:

Example 7

(1) Consider the following two-player coordination game:

	L	R
T	1,0	0,2
B	0,3	2,-1

Suppose that each player knows nothing about his opponent's strategy. The row player chooses a distribution $(p, 1-p)$ over his set of actions $\{T, B\}$. When playing $(p, 1-p)$, the row player's expected payoff is $\min(p, 2(1-p))$. The maximum over p is achieved when $p = \frac{2}{3}$. The same reasoning applies to player 2: she plays $(\frac{1}{2}, \frac{1}{2})$ and the only partially-specified equilibrium, when both players' information is trivial, is $((\frac{2}{3}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{2}))$. Note that L receives a positive probability, although it is not a best response to $(\frac{2}{3}, \frac{1}{3})$ (whether or not player 2 knows about this strategy). Furthermore, when player 2's knowledge of the strategy played by player 1 is trivial, R is the worst strategy that player 2 can play against $(\frac{2}{3}, \frac{1}{3})$. Nevertheless, R is played in equilibrium with a positive probability.

(2) Consider the following two-player game:

	L	M	R
T	3,3	0,0	0,0
C	0,0	2,2	0,0
B	0,0	0,0	1,1

Consider the case where player 1 knows the probability assigned by player 2 to R and player 2 knows the probability assigned by player 1 to B. In other words, the sub-algebra of events whose probabilities are known to player 1, $\mathcal{A}_1$, is generated by the partition $\{\{L, M\}, \{R\}\}$ and the corresponding sub-algebra known to player 2, $\mathcal{A}_2$, is generated by the partition $\{\{T, C\}, \{B\}\}$. In equilibrium, player 1 plays (p_1, p_2, p_3) and player 2 plays (q_1, q_2, q_3). When $p_1 + p_2 > 0$, as in example (1), $\frac{p_1}{p_2} = \frac{q_1}{q_2} = \frac{2}{3}$.

There are three equilibria: a. $p = (\frac{2}{5}, \frac{3}{5}, 0)$, $q = (\frac{2}{5}, \frac{3}{5}, 0)$; b. $p = (0, 0, 1)$, $q = (0, 0, 1)$; and c. $p = (\frac{2}{11}, \frac{3}{11}, \frac{6}{11})$, $q = (\frac{2}{11}, \frac{3}{11}, \frac{6}{11})$.

(4) In examples (1)-(2) each player knows the probabilities of some subsets of the other player's set of strategies and these subsets form an algebra. In the following example, the set of subsets whose probability is known to player 2 is not an algebra.

Consider the following zero-sum game:

	s_1	s_2	s_3	s_4
T	0	1	0	-1
R	1	0	1	2

Suppose that player 2 knows the mixed strategy played by player 1. Player 1, on the other hand, is informed of the probability of $A = \{s_1, s_2\}$ and $B = \{s_2, s_3\}$ (and, as usual, of $S = \{s_1, s_2, s_3, s_4\}$). There is no equilibrium where one of the players plays a pure strategy. Thus, player 1 must play $(\frac{1}{2}, \frac{1}{2})$, in which case any mixed strategy of player 2 is a best response.

If player 1 plays $(\alpha, 1-\alpha)$, then the payoff matrix reduces to

	s_1	s_2	s_3	s_4
	$1-\alpha$	α	$1-\alpha$	$2-3\alpha$

Suppose that player 2 plays the mixed strategy P. Player 1 evaluates his strategies using the partially-specified probability $(P, \mathcal{Y})$, where $\mathcal{Y} = \{A, B, S\}$.[9] This is a similar situation to that discussed in Example 4. The expected payoff of player 1 is precisely $\int \psi d_{\mathcal{Y}} P$, with ψ being the function analyzed in Example 4. When $\alpha \leq \frac{1}{2}$, this integral was already found to be $\max((1-\alpha)P(S \setminus A) + (1-\alpha)P(S \setminus B) - \alpha, \alpha P(A) + (1-\alpha)P(S \setminus A))$. If this maximum is attained in the left element, $(1-\alpha)P(S \setminus A) + (1-\alpha)P(S \setminus B) - \alpha$, then player 1 will choose $\alpha = 0$ in order to maximize his payoff. However, this is impossible. Thus, the maximum is attained in the right element, $\alpha P(A) + (1-\alpha)P(S \setminus A)$. The maximum can be obtained at $\alpha = \frac{1}{2}$ only if $P(S \setminus A) = P(A) = \frac{1}{2}$ in which case the expected payoff is $\frac{1}{2}$.

A similar calculation is conducted for $\alpha \geq \frac{1}{2}$. In this case, $\alpha \geq 1-\alpha \geq 2-3\alpha$. Therefore, $\int \psi d_{\mathcal{Y}} P = \max((1-\alpha)P(A) + (1-\alpha)P(B) - (2-3\alpha), (1-\alpha)P(A) + (2-3\alpha)P(S \setminus A))$. The maximum is achieved at $\alpha = \frac{1}{2}$. Thus, in a partially-specified equilibrium player 1 plays $(\frac{1}{2}, \frac{1}{2})$ and player 2 assigns probability $\frac{1}{2}$ to the strategies in A. In other words, any mixed strategy of player 2 that assigns probability $\frac{1}{2}$ to A forms (together with player 1's strategy of $(\frac{1}{2}, \frac{1}{2})$) a partially-specified equilibrium. ■

The following theorem ensures the existence of a partially-specified equilibrium for every information structure:

Theorem 2 *Let $\mathcal{Y}_i^j$ be the sets of random variables whose expectations are known to player i, $i \in M$ ($j \neq i$). Then, there exists a partially-specified equilibrium w.r.t. $\mathcal{Y}_i^j$ ($j \neq i$).*

This theorem can be proven by a standard fixed-point technique that is based on two facts: that the set of player i's strategies which are best responses to any p_{-i} is convex and that the best-response correspondence is upper semi-continuous. The details are omitted.

In Nash equilibrium, each player plays a best response to the actual strategies of the other players. In other words, Nash equilibrium requires that beliefs be rationalized by actual behavior. In partially-specified equilibrium, players play their best response to the partial specification they obtain regarding the actual strategies of the other players.

Dow and Werlang (1994) defined equilibrium with non-additive probabilities in which players play their best response according to their beliefs. However, aside from one restriction, namely that only strategies which appear in the set defined to be the support[10]

[9] Here an event is identified with its characteristic function.

[10] The support of a non-additive capacity has several plausible interpretations.

are allowed to be played, there is no relation in this equilibrium between the strategies played and the beliefs. Such a degree of latitude seems to allow for an excessive degree of freedom. As a result, players' beliefs are essentially unrelated to their actual behavior. Eichberger and Kelsey (1999) extended the definition of Dow and Werlang (1994) from two to n-player games. Marinacci (2000) also defined equilibrium with non-additive probabilities, but required different conditions of consistency between the beliefs and the actual strategies played.

Klibanoff (1996) and Lo (1999) defined equilibrium notions in which players' beliefs about others is represented by a set of mixed strategies. However, since there is no intrinsic connection between players' beliefs and what others actually do, some consistency between the two is called for. This kind of requirement is not required here.

VIII. Discussion

A. Ambiguity aversion - I

Ghirardato and Marinacci (2002) compare between different preference orders from the ambiguity aversion perspective. They introduce a notion of 'being more ambiguity averse than', and show that one maxmin utility maximization model is more ambiguity averse than another, if and only if its corresponding set of priors is included in that corresponding to the other. A natural question arises as to what this notion means in the present model.

Fix a utility function u and denote by $\succsim_i$ the preference order induced by u and the partially-specified probability $(P_i, \mathcal{Y}_i)$, $i = 1, 2$. The result of Ghirardato and Marinacci (2002) implies that $\succsim_2$ is more ambiguity averse than $\succsim_1$, if and only if $\mathcal{Q}(P_1, \mathcal{Y}_1) \subseteq \mathcal{Q}(P_2, \mathcal{Y}_2)$. This statement characterizes when one preference order is more ambiguity averse than another in terms of their corresponding sets of priors. However, when the decision is based on partially-specified probabilities, it can be phrased in terms of the information made available to the decision maker.

Recall the notation $\text{span}(\mathcal{Y}_i)$ introduced in Section V, and that the decision maker knows not only the expected value of any function in $\mathcal{Y}_i$, but also of those in $\text{span}(\mathcal{Y}_i)$.

It turns out that $\mathcal{Q}(P_1, \mathcal{Y}_1) \subseteq \mathcal{Q}(P_2, \mathcal{Y}_2)$ if and only if (i) $\text{span}(\mathcal{Y}_2) \subseteq \text{span}(\mathcal{Y}_1)$; and (ii) P_1 and P_2 agree the variables in $\text{span}(\mathcal{Y}_2)$. These two conditions say that the set $\text{span}(\mathcal{Y}_1)$, about which the decision maker is informed by $(P_1, \mathcal{Y}_1)$, is larger than the set that she is informed about by $(P_2, \mathcal{Y}_2)$. Moreover, P_1 and P_2 agree on the smaller set. This implies that the information about the underlying probability becomes more specific under $(P_1, \mathcal{Y}_1)$: the set of priors it induces is smaller than that induced by $(P_2, \mathcal{Y}_2)$.

B. Ambiguity aversion - II

This analysis introduces a first-order approximation to the manner in which people make decisions in the presence of a partially-specified probability. In Ellsberg's original decision problem, Gamble **X** is weakly dominated by Gamble **Z**. Nevertheless, the theory of decision making with a partially-specified probability would predict that **X** and **Z** are equivalent. To make things even worse, suppose that **X** is modified somewhat so that instead of \$100, the prize for drawing a red ball is \$101. In this case, **X** is predicted to be strictly preferred to **Z**.

The reason for this difficulty is that the theory takes an extreme ambiguity-aversion approach whereby any information provided is accepted without question and everything else is ignored.

Similar difficulties arise with expected utility theory. A situation may occur whereby one act weakly dominates another, even though they are actually equivalent. This can happen when large prizes are ascribed a probability of zero and are therefore not taken into account in expected utility.

To improve upon the current theory, one might consider discriminating between information sources according to their reliability. More reliable sources would receive a greater weight than less reliable ones. In this case, wild guesses would also be reliable to a certain extent and would therefore be taken into account with a weight determined according to their level of reliability.

Recall eq. (1). In order to formalize what was stated in the previous paragraph, let $\mathcal{Y}$ be a set of random variables and suppose that v is a real function defined on $\mathcal{Y}$. $v(Y)$ is interpreted as 'the expectation of Y is claimed to be $v(Y)$' for every $Y \in \mathcal{Y}$. However, the function v can summarize the information received from various sources. The reliability of these sources might vary.

One could think of a reliability factor r_Y attached to every $Y \in \mathcal{Y}$. This factor is meant to indicate the extent to which the information about Y is reliable. It can then be taken into account when evaluating a function ψ, as follows:

$$\int \psi dP_{\mathcal{Y}} = \max\Big\{ \sum_{Y\in\mathcal{Y}} r_Y \lambda_Y v(Y);\ \sum_{Y\in\mathcal{Y}} \lambda_Y Y \le \psi \text{ and } \lambda_Y \ge 0 \text{ for every } Y \in \mathcal{Y} \Big\}.$$

Note that $\lambda_Y v(Y)$ is discounted by the coefficient r_Y to obtain $r_Y \lambda_Y v(Y)$.

C. On the definition of SFaF

The definition of a strongly fat-free axiom makes use of the constant act $\mathbf{c}_1$. Based on this definition Axiom (iii) is formulated. This is precisely the axiom needed in order to determine the existence of vN-M representation on L_c, which in turn guarantees the existence of a maximal element $\mathbf{M}$ in L_c. Lemma 4 states that the definition of SFaF could have used $\mathbf{M}$ rather than $\mathbf{c}_1$. However, at that point the existence of $\mathbf{M}$ is not yet guaranteed.

Instead of this definition of SFaF, one could avoid using $\mathbf{c}_1$ by adopting a more stringent definition in which f is a strongly fat-free act if any combination of f with *any* constant act is FaF. In the presence of the other axioms, this definition would be equivalent to the above version.

D. On the continuity axiom

The continuity axiom used above is a bit stronger than that used by Anscombe and Aumann (1963). It was required in (iv)(a) (the first part of the continuity axiom) that if $f \succ g$ and $g \succsim h$, then there is an α in $(0, 1)$ such that $\alpha f + (1 - \alpha)h \succ g$. The vN-M axiom requires that the same conclusion holds under a weaker condition, i.e. $f \succ g$ and $g \succ h$.

The stronger axiom is needed in the proof of Lemmas 4 and 5. The independence axiom holds only for strong fat-free acts. A priori, there is only one fat-free act: $\mathbf{m}$ (if it exists). Axioms (i), (iv) and a weaker version of (iii) can only ensure that there is a vN-M utility representation of $\succsim$ over L_c. However, to ensure a broader scope for the independence axiom, it is essential, as guaranteed by Lemma 5, that for every act there is an equivalent act which is fat-free. A weaker version of the continuity axiom would not be sufficient for this purpose.

E. Additivity and fat-free acts

It turns out that when the probability is specified on a sub-algebra of events,[11] the integral in eq. (1) is additive over the set of functions that are measurable with respect to this sub-algebra. Therefore, the implied expected utility is additive when restricted to fat-free acts measurable w.r.t. at least one function in that set of functions. This phenomenon is strongly related to the model of Epstein and Zhang (2001) in which the probability restricted to the set of unambiguous events is additive. In that case, the set is more general than an algebra since it can be a λ-system.

[11] That is, the expected values, of indicator functions of events in a sub-algebra, are given

When the probability is partially specified (not over a sub-algebra), additivity is preserved over the strongly fat-free acts, which by Lemma 6 form a convex set. This does not imply that a convex combination of (not strongly) fat-free acts is necessarily fat-free.

F. The Bayesian approach and partially-specified probabilities

An orthodox Bayesian approach would dictate that whenever only partial information about the probability is available, the decision maker adopts an additive probability that is consistent with the data. For instance, in Ellsberg's urn, since only the probability of white or black is known while the probability of each color separately is not, a Bayesian decision maker should assign a probability to white and to black. A uniform distribution, namely probability of $\frac{1}{3}$ to both white and black, seems 'natural' in this relatively simple case.

Suppose, however, that there are five states of nature: $S = \{s_1, s_2, s_3, s_4, s_5\}$. Furthermore, according to the information available, the probability of $\{s_1, s_2, s_3\}$ and of $\{s_3, s_4\}$ is .9 for each. Obviously, a uniform distribution over $\{s_1, s_2, s_3\}$ or over $\{s_3, s_4\}$ is inconsistent with any distribution that assigns a probability of .9 to $\{s_1, s_2, s_3\}$ and to $\{s_3, s_4\}$. A reasonable option would be to adopt the distribution that maximizes the entropy of all probability distributions that are consistent with the information available. This would be a natural alternative to the uniform in cases where such distribution is impossibility.

Entropy has some desirable properties that become important when applied to measuring the quantity of information. However, when it comes to decision-making problems, entropy has no greater role than any other (symmetric) concave function. This assertion is bolstered by Blackwell (1953) who compared information structures in relation to utility maximization. One of his results refers to *all* concave functions rather than to any particular one.

IX. Final comments

A. Partially-specified probability and the Choquet integral

Assume a simple case of a partially-specified probability, where $\mathcal{Y}$ contains only the indicator functions of all events in some algebra $\mathcal{A}$. In such a case we denote the partially-specified probability by $(P, \mathcal{A})$.

Remark 2 (a) If ψ is measurable with respect to $\mathcal{A}$, then this integral coincides with the regular expectation.

(b) Let $(P, \mathcal{A})$ be a probability specified on a sub-algebra. Since S is finite, $\mathcal{A}$ is generated by a partition, say $\mathcal{P}$, of S. Thus, the integral can also be written with the further restriction that all sets E are taken from the partition $\mathcal{P}$ and, moreover, that the coefficients λ_E are non-negative:

$$(5) \qquad \int \psi dP_{\mathcal{A}} = \max\Big\{ \sum_{E\in\mathcal{P}} \lambda_E P(E);\ \sum_{E\in\mathcal{P}} \lambda_E \mathbb{I}_E \leq \psi \text{ and } \lambda_E \geq 0 \text{ for every } E \in \mathcal{P}\Big\}.$$

■

A probability specified on a sub-algebra, $(P, \mathcal{A})$, induces a capacity $v_{(P,\mathcal{A})}$ defined over S: $v_{(P,\mathcal{A})}(E) = \int \mathbb{I}_E dP_{\mathcal{A}}$ which is convex.[12] By Remark 2(b), $v_{(P,\mathcal{A})} = \max_{F\subseteq E,\ F\in\mathcal{A}} P(F)$ and, according to Lehrer (2009), the Choquet integral w.r.t. $v_{(P,\mathcal{A})}$ coincides with the integral in eq. (5). In other words, the integral with respect to a probability specified on a sub-algebra $(P, \mathcal{A})$ coincides with the Choquet integral w.r.t. $v_{(P,\mathcal{A})}$.

The conclusion above does not hold in the general case. Let $(P, \mathcal{Y})$ be a partially-specified probability and define the capacity $v_{P,\mathcal{Y}}$ in a manner similar to $v_{P,\mathcal{A}}$. Teper (2009) shows an example in which the integral w.r.t. $(P, \mathcal{Y})$ does not coincide with the Choquet integral w.r.t. $v_{(P,\mathcal{A})}$. Furthermore, it is easy to construct an example and verify that the integral w.r.t. a partially specified probability is, in general, not comonotonic additive, and thus does not coincide with the Choquet integral.

B. Convex capacities and partially-specified probabilities

Let v be a convex capacity. An event E in S is called *fat-free* if $F \subsetneqq E$ implies $v(F) < v(E)$. Let $(P, \mathcal{A})$ be a probability specified on a sub-algebra and define the capacity $v_{P,\mathcal{A}}$ as described above. It is clear that $v_{P,\mathcal{A}}$ is convex (see footnote 12). The following proposition characterizes those convex capacities that are of the form $v_{P,\mathcal{A}}(E)$.

Proposition 2 *Let v be a convex non-additive probability. There exists a probability specified on a sub-algebra $(P, \mathcal{A})$ such that $v = v_{P,\mathcal{A}}$ if and only if for every fat-free event E and for every F, $v(E) + v(F) = v(E \cup F) + v(E \cap F)$.*

The proof appears in the Appendix.

Let $T \subseteq S$. A *unanimity capacity* u_T is defined as $u_T(E) = 1$ if $T \subseteq E$ and $u_T(E) = 0$ otherwise. It turns out that a unanimity capacity is also of the form $v_{P,\mathcal{A}}$: $\mathcal{A} = \{T, S \setminus T\}$,

[12] A capacity v is *convex* if for every $E, F \subseteq S$, $v(E) + v(F) \leq v(E \cup F) + v(E \cap F)$.

$P(T) = 1$ and $P(S \backslash T) = 0$. Moreover, a capacity $v_{P,\mathcal{A}}$ is a convex combination of unanimity games of a special kind, as demonstrated by the following lemma (in which a sub-algebra $\mathcal{A}$ generated by the partition $\mathcal{Q}$ is denoted by $\mathcal{Q}(\mathcal{A})$):

Lemma 8 *Let v be a capacity. There is a probability specified on a sub-algebra, $(P, \mathcal{A})$, such that $v = v_{P,\mathcal{A}}$ if and only if there is a partition $\mathcal{Q}$ of S such that $v = \sum_{T \in \mathcal{Q}} \alpha_T u_T$, where $\alpha_T \geq 0$ and $\sum_{T \in \mathcal{Q}} \alpha_T = 1$. Moreover, $\mathcal{Q} = \mathcal{Q}(\mathcal{A})$.*

The proof is straightforward and therefore omitted.

References

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

[2] Aumann, Robert J. (1974) "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, **1**, 67-96.

[3] Aumann, Robert J. (1987) "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica,* **55**, 1-18.
Azrieli, Yaron and Ehud Lehrer (2007a) "Extendable cooperative games," *Journal of Public Economic Theory*, **9**, 1069-1078

[4] Azrieli, Yaron and Ehud Lehrer (2007b) "On some families of cooperative fuzzy games," *International Journal of game Theory*, **36**, 1-15.

[5] Bewley, Truman F. (2002) "Knightian decision theory. Part I," *Decisions in Economics and Finance* **25**, 79-110.

[6] Blackwell, David (1953) "Equivalent comparison of experiments," *Annals of Mathematical Statistics*, **24**, 265-272.

[7] de Finetti, Bruno (1970) *Teoria Delle Probabilità*. Einaudi, Turin. An English translation by Antonio Machi and Adrian Smith: Theory of Probability, Wiley, London, two volumes, 1974 and 1975.

[8] Dempster, Arthur P. (1967) "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics* **38**, 325-339

[9] Dow, James and Sergio Werlang (1994) "Nash Equilibrium under Knightian uncertainty: breaking down backward induction, *Journal of Economic Theory*, **64**, 305-324.

[10] Eichberger, Jurgen and David Kelsey (1999) "Non-Additive beliefs and strategic equilibria," *Games and Economic Behavior*, **30**, 183-215.

[11] Epstein, Larry G. and Jiankang Zhang (2001) "Subjective probabilities on subjectively unambiguous events," *Econometrica*, **69**, 265-306.

[12] Ghirardato, Paolo and Massimo Marinacci (2002)"Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory* **102**, 251-289.

[13] Gilboa, Itzhak and David Schmeidler (1989) "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics,* **18**, 141-153.

[14] Klibanoff, Peter (1996) "Uncertainty, decision, and normal form games," mimeo, Northewestern University

[15] Lehrer, Ehud (2009) "A new integral for capacities," *Economic Theory*, **39**, 157-176.

[16] Lo, Kin C. (1999) "Extensive Form Games with Uncertainty Averse Players," *Games and Economic Behavior*, **28**, 256-270.

[17] Maccheroni, Fabio, Massimo Marinacci and Aldo Rustichini (2006), "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, **74**, 1447-1498.

[18] Machina, Mark J. (1989) "Dynamic consistency and non-expected utility models of choice under uncertainty," *Journal of Economic Literature*, **27**, 1622-1688.

[19] Marinacci, Massimo (2000) "Ambiguous Games," *Games and Economic Behavior*, **31**, 191-219.

[20] von Neumann, John and Oskar Morgenstern (1944) *Theory of games and economic behavior*, Princeton University Press.

[21] Savage, Leonard J. (1954) *The Foundations of statistics*. New York, NY: Wiley.

[22] Schmeidler, David (1989) "Subjective probabilities without additivity," *Econometrica*, **57**, 571-587.

[23] Shafer, Glen (1976) *A Mathematical Theory of Evidence*. Princeton University Press

[24] Teper, Roee (2009) "Time continuity and non-additive expecetd utility," *Mathematics of Operations Research*, **34**, 661-673.

[25] Zhang, Jiankang (2002) "Subjective, ambiguity, expected utility and Choquet expected utility," *Economic Theory* **20**, 159181.

Appendix

Proof of Lemma 1. Since $\mathcal{Y}$ has a finite dimension, there is a finite set, $\mathcal{Y}'$, such that span$(\mathcal{Y}') =$ span$(\mathcal{Y})$. Since $\mathcal{Y}' \subseteq \mathcal{Y}$, $\int \psi dP_{\mathcal{Y}'} \leq \int \psi dP_{\mathcal{Y}}$ for every ψ.

By definition, $\int \psi dP_{\mathcal{Y}} = \sum_{Y \in \mathcal{Y}} \lambda_Y E_P(Y)$, where $\sum_{Y \in \mathcal{Y}} \lambda_Y Y \leq \psi$. Since $Y \in \mathcal{Y}$, there are coefficients θ_Y^Z, $Z \in \mathcal{Y}'$, such that $Y = \sum_{Z \in \mathcal{Y}'} \theta_Y^Z Z$. Since E_P is additive over $\mathcal{Y}'$, $\sum_{Y \in \mathcal{Y}'} \lambda_Y E_P(Y) = \sum_{Y \in \mathcal{Y}} \lambda_Y E_P(\sum_{Z \in \mathcal{Y}'} \theta_Y^Z Z) = \sum_{Y \in \mathcal{Y}} \sum_{Z \in \mathcal{Y}'} \lambda_Y \theta_Y^Z E_P(Z)$. The last summation is a linear combination of elements from $\mathcal{Y}'$. This implies that $\int \psi dP_{\mathcal{Y}} \leq \int \psi dP_{\mathcal{Y}'}$ and equality is established. ■

Proof of Lemma 2. Let $\int \psi dP_{\mathcal{Y}} = \sum_{Y\in\mathcal{Y}} \lambda_Y P(Y)$, where $\sum_{Y\in\mathcal{Y}} \lambda_Y Y \leq \psi$ and $\int \phi dP_{\mathcal{Y}} = \sum_{Y\in\mathcal{Y}} \gamma_Y P(Y)$, where $\sum_{Y\in\mathcal{Y}} \gamma_Y Y \leq \phi$. Thus, $\alpha \sum_{Y\in\mathcal{Y}} \lambda_Y Y + (1-\alpha)\sum_{Y\in\mathcal{Y}} \gamma_Y Y \leq \alpha\psi + (1-\alpha)\phi$. The left-hand side is one of the summations in the right-hand side of eq. (1) and therefore the desired result is obtained. ■

Proof of Lemma 3. By Lemma 1 we can assume that $\mathcal{Y}$ is finite. Denote by Δ the set of non-negative functions ψ over S such that $\sum_{s\in S} \psi(s) = 1$ and $D = \text{conv}(\Delta \cup -\Delta)$. Since $\int \bullet\, dP_{\mathcal{Y}}$ is homogenous (as a function defined on $[0,1]^S$), we can assume that $\mathcal{Y} \subseteq \Delta$.

Define the function T over D as the least concave function on D that satisfies $T(\psi) \geq P(\psi)$ and $T(-\psi) \geq -P(\psi)$ for every $\psi \in \mathcal{Y}$. By Lemma 2, $\int \bullet\, dP_{\mathcal{Y}}$ is a concave function over Δ and due to homogeneity it is a concave function over all $[0,1]^S$. Thus, $T(\psi) = \int \psi dP_{\mathcal{Y}}$ for every non-negative $\psi \in D$.

Since $\mathcal{Y}$ is finite, T is piecewise linear. Thus, D can be split into a finite number of closed sets, on each of which T is linear. As a piecewise linear concave function, T can also be expressed as a minimum of **finitely** many affine functions. That is, there are finitely many S-dimensional vectors Q_i and scalars b_i such that[13] $T(\psi) = \min_i \psi \cdot Q_i + b_i$ for every $\psi \in D$. Since $\sum_{s\in S} \psi(s) = 1$ for every $\psi \in \Delta$, one can find S-dimensional vectors Q'_i such that $\min_i \psi \cdot Q_i + b_i = \psi \cdot Q'_i$. Therefore, T is a minimum of finitely many linear functions over Δ. Since T is non-negative over Δ, these vectors are non-negative.

It remains to show that the vectors Q'_i are all probability vectors. Suppose, on the contrary, that there exists $\psi \in \Delta$ such that there is no probability vector P for which $P \cdot \psi = T(\psi)$ and for any other $\phi \in \Delta$, $P \cdot \phi \leq T(\phi)$. This implies that there is $0 < a < 1$ such that $T(\alpha\psi + (1-\alpha)\mathbb{I}_S) > \alpha T(\psi) + (1-\alpha)T(\mathbb{I}_S)$. However, $T(\alpha\psi + (1-\alpha)\mathbb{I}_S) = \int \alpha\psi + (1-\alpha)\mathbb{I}_S dP_{\mathcal{Y}} = \int \alpha\psi dP_{\mathcal{Y}} + \int (1-\alpha)\mathbb{I}_S dP_{\mathcal{Y}} = \alpha \int \psi dP_{\mathcal{Y}} + (1-\alpha)\int \mathbb{I}_S dP_{\mathcal{Y}} = \alpha T(\psi) + (1-\alpha)T(\mathbb{I}_S)$. This creates a contradiction. We conclude that the integral of a partially-specified probability is a minimum of finitely many regular (additive) integrals. ■

Proof of Lemma 4: If f is SFaF, then for every $\alpha \in [0,1)$, $\alpha f + (1-\alpha)\mathbf{c}_1$ is FaF. Let h be such that $h < \alpha f + (1-\alpha)\mathbf{M}$. We show that $h \prec \alpha f + (1-\alpha)\mathbf{M}$ and thereby prove that $\alpha f + (1-\alpha)\mathbf{M}$ is SFaF.

There exists $0 < \beta \leq 1$ such that $\beta\mathbf{M} + (1-\beta)\mathbf{m} \sim \mathbf{c}_1$. We can assume w.l.o.g. that $\beta < 1$. Define $\gamma = \frac{\frac{\beta}{1-\beta}}{1-\alpha+\frac{\beta}{1-\beta}}$. Since $h < \alpha f + (1-\alpha)\mathbf{M}$, $\gamma h + (1-\gamma)\mathbf{m} < \gamma\alpha f + \gamma(1-\alpha)\mathbf{M} + (1-\gamma)\mathbf{m} = \gamma\alpha f + (1-\gamma\alpha)\mathbf{c}_1$. Since f is SFaF, $\gamma h + (1-\gamma)\mathbf{m} \prec \gamma\alpha f + (1-\gamma\alpha)\mathbf{c}_1$. Moreover, by (vi), $\mathbf{m}$ is SFaF and therefore, due to (iii), $\gamma h + (1-\gamma)\mathbf{m} \prec \gamma\alpha f + (1-\gamma\alpha)\mathbf{c}_1 = \gamma\alpha f + \gamma(1-\alpha)\mathbf{M} + (1-\gamma)\mathbf{m}$ implies (since $\gamma > 0$) that $h \prec \alpha f + (1-\alpha)\mathbf{M}$, as desired.

As for the opposite direction, assume that $\beta f + (1-\beta)\mathbf{M}$ is FaF for every $\beta \in (0,1)$ and that $h < \alpha f + (1-\alpha)\mathbf{c}_1$. We show that $h \prec \alpha f + (1-\alpha)\mathbf{c}_1$. As before, $\beta\mathbf{M} + (1-\beta)\mathbf{m} \sim \mathbf{c}_1$, with $0 \leq \beta \leq 1$. There exists an act g that satisfies, $h(s) \sim \Big((1-(1-\alpha)(1-\beta))g + (1-\alpha)(1-\beta)\mathbf{m}\Big)(s)$ for every $s \in S$. This is so because $\Big((1-(1-\alpha)(1-\beta))\mathbf{m} + (1-\alpha)(1-\beta)\mathbf{m}\Big)(s) \leq h(s) \leq \Big((1-(1-\alpha)(1-\beta))(\frac{\alpha}{(1-(1-\alpha)(1-\beta)}f + \frac{(1-\alpha)\beta}{(1-(1-\alpha)(1-\beta)}\mathbf{M}) + (1-\alpha)(1-\beta)\mathbf{m}\Big)(s) \leq \Big((1-(1-\alpha)(1-\beta))\mathbf{M} + (1-\alpha)(1-\beta)\mathbf{m}\Big)(s)$. Thus, $(1-(1-\alpha)(1-\beta))g + (1-\alpha)(1-\beta)\mathbf{m} < \alpha f + (1-\alpha)(\beta\mathbf{M} + (1-\beta)\mathbf{m}) = \alpha f + (1-\alpha)\beta\mathbf{M} + (1-\beta)\mathbf{m}$. Moreover, $g < \frac{\alpha}{1-(1-\alpha)(1-\beta)}f + \frac{(1-\alpha)\beta}{1-(1-\alpha)(1-\beta)}\mathbf{M}$. By assumption,

[13] Here, '$\cdot$' denotes the inner product between two vectors.

$g \prec \frac{\alpha}{1-(1-\alpha)(1-\beta)} f + \frac{(1-\alpha)\beta}{1-(1-\alpha)(1-\beta)}\mathbf{M}$. Since $\mathbf{m}$ is SFaF and using (iii), $(1-(1-\alpha)(1-\beta))g + (1-\alpha)(1-\beta)\mathbf{m} \prec (1-(1-\alpha)(1-\beta))(\frac{\alpha}{1-(1-\alpha)(1-\beta)} f + \frac{(1-\alpha)\beta}{1-(1-\alpha)(1-\beta)}\mathbf{M}) + (1-\alpha)(1-\beta)\mathbf{m} = \alpha f + (1-\alpha)\mathbf{c}_1$. The left-hand side is equal to h and thus, $h \prec \alpha f + (1-\alpha)\mathbf{c}_1$, as desired. ■

Proof of Lemma 5. First, note that f itself is a member of the set $A(f) = W(f) \cap \{g;\ f \sim g\}$. The independence axiom (iii) applies in particular to L_c, and along with (i) and (vi) implies that $\succsim$ is continuous over L_c. Using a diagonalization method, for instance, one can find an infimum, say h, w.r.t. $\geq$ in $A(f)$. The act h satisfies:
(a) $h \not\geq g$ for every $g \in A(f)$; and
(b) if h' satisfies $h' \geq h$ and $h'(s) \succ h(s)$ whenever $f(s) \succ h(s)$, then there exists $g' \in A(f)$ such that $h' > g'$.

We show that $h = [f]$. Suppose on the contrary, that $f \succ h$. By (ii) $f > h$ and therefore, $\beta f + (1-\beta)h > h$ for every $\beta \in [0,1)$. From (b) it follows that for every $\beta \in [0,1)$ there is $g'(\beta) \in A(f)$ such that $\beta f + (1-\beta)h > g'(\beta)$.

By (iv), there exists $\alpha \in (0,1)$ such that $f \succ \alpha f + (1-\alpha)h$. This is a contradiction because by (ii), $\alpha f + (1-\alpha)h \succsim g'(\alpha) \sim f$. We thus obtained that $f \sim h$ and h is FaF, as desired. ■

Proof of Lemma 6: Assume that f and g are SFaF and let $0 < \alpha < 1$. Let $0 \leq \beta < 1$ and assume an act h such that $h < \beta(\alpha f + (1-\alpha)g) + (1-\beta)\mathbf{M}$. We show that $h \prec \beta(\alpha f + (1-\alpha)g) + (1-\beta)\mathbf{M}$. Define the act h': for every $s \in S$, $h'(s) = \max\Big((\beta\alpha f + (1-\beta\alpha)\mathbf{m})(s), h(s)\Big)$, where the maximum is taken w.r.t. $\succ$.

By definition, $h \leq h'$. By (ii), $h \preceq h'$. Moreover, $h' < \beta\alpha f + (1-\beta\alpha)\Big(\frac{\beta(1-\alpha)}{1-\beta\alpha} g + \frac{(1-\beta)}{1-\beta\alpha}\mathbf{M}\Big)$. Thus, there exists an act, say h'', such that $h' = \beta\alpha f + (1-\beta\alpha)h''$. Therefore, $h'' < \frac{\beta(1-\alpha)}{1-\beta\alpha} g + \frac{(1-\beta)}{1-\beta\alpha}\mathbf{M}$. Since g is SFaF, $h'' \prec \frac{\beta(1-\alpha)}{1-\beta\alpha} g + \frac{(1-\beta)}{1-\beta\alpha}\mathbf{M}$.

The act f is also SFaF and therefore by (iii), $\beta\alpha f + (1-\beta\alpha)h'' \prec \beta\alpha f + (1-\beta\alpha)\Big(\frac{\beta(1-\alpha)}{1-\beta\alpha} g + \frac{(1-\beta)}{1-\beta\alpha}\mathbf{M}\Big) = \beta(\alpha f + (1-\alpha)g) + (1-\beta)\mathbf{M}$. The left-hand side is h' and therefore this is the desired result. ■

Proof of Lemma 7: Suppose, contrary to the lemma, that there is a sequence α_n which converges to zero and that $[\alpha_n f + (1-\alpha_n)\mathbf{M}]$ is not SFaF. Thus, for every n, there is $0 \leq \beta_n < 1$ and an act g_n such that $g_n < \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M}$ and $g_n \sim \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M}$.

There is no act h_n such that $g_n = \beta_n h_n + (1-\beta_n)\mathbf{M}$. Indeed, if such an h_n exists, then $\beta_n h_n + (1-\beta_n)\mathbf{M} < \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M}$ and $\beta_n h_n + (1-\beta_n)\mathbf{M} \sim \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M}$. Since $\mathbf{M}$ is SFaF and by (iii), this implies that $h_n \sim [\alpha_n f + (1-\alpha_n)\mathbf{M}]$ and $h_n < [\alpha_n f + (1-\alpha_n)\mathbf{M}]$, contrary to the definition of $[\alpha_n f + (1-\alpha_n)\mathbf{M}]$. Let γ be the smallest constant such that there exists h_n whereby $g_n = \gamma_n h_n + (1-\gamma_n)\mathbf{M}$ (such γ_n and h_n exist since $\succsim$ has a vN-M representation on the set $\{\ell;\ g_n$ is a convex combination of ℓ and $\mathbf{M}\}$). It must be that $\gamma_n > \beta_n$ and that, for at least one $s \in S$, $g_n(s) = \mathbf{m}(s)$.

We obtain that $\gamma_n h_n + (1-\gamma_n)\mathbf{M} < \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M} \leq \beta_n\alpha_n f + (1-\beta_n\alpha_n)\mathbf{M}$ and $\gamma_n h_n + (1-\gamma_n)\mathbf{M} \sim \beta_n[\alpha_n f + (1-\alpha_n)\mathbf{M}] + (1-\beta_n)\mathbf{M} \sim \beta_n\alpha_n f + (1-\beta_n\alpha_n)\mathbf{M}$. The latter implies that $h_n \sim \frac{\beta_n\alpha_n}{\gamma_n} f + \frac{(\gamma_n - \beta_n\alpha_n)}{\gamma_n}\mathbf{M}$.

Since there are finitely many states, there are infinitely many n's and a state s_0 such that $h_n(s_0) = \mathbf{m}(s_0)$. Notice that the act h, which coincides with $\mathbf{M}$ on $S \setminus \{s_0\}$ and with $\mathbf{m}$ on s_0, satisfies $h_n \leq h$ for infinitely many n's. Thus, by (ii), $h \succsim \frac{\beta_n \alpha_n}{\gamma_n} f + \frac{(\gamma_n - \beta_n \alpha_n)}{\gamma_n}\mathbf{M}$ infinitely often. Since $\frac{(\gamma_n - \beta_n \alpha_n)}{\gamma_n}$ tends to 0, $h \sim \mathbf{M}$, which means that $\mathbf{M}$ is not SFaF, contradicting (v). ■

Proof of Proposition 1. Let $\mathcal{Y}$ be the set of SFaF functions. This set is not empty (by (5)). Consider the set $D = \{\sum_{X \in \mathcal{Y}} \alpha_X (X - I(X)\mathbb{I}_S);\ \alpha_X \in \mathbb{R}\}$. D is convex and is disjoint from the open negative orthant. Indeed, if there exists a summation $\sum_{X \in \mathcal{Y}} \alpha_X (X - I(X)\mathbb{I}_S) = Z << 0$ (which implies that D and the open negative orthant are not disjoint), then $\sum_{\alpha_X > 0} \alpha_X X + \sum_{\alpha_X < 0} |\alpha_X| I(X)\mathbb{I}_S << \sum_{\alpha_X < 0} |\alpha_X| X + \sum_{\alpha_X > 0} \alpha_X I(X)\mathbb{I}_S$. By (1), (3), (4) and (6), $\sum_{\alpha_X > 0} \alpha_X I(X) + \sum_{\alpha_X < 0} |\alpha_X| I(X) < \sum_{\alpha_X < 0} |\alpha_X| I(X) + \sum_{\alpha_X > 0} \alpha_X I(X)$, which is a contradiction.

The separation theorem ensures the existence of a non-negative vector $P \neq 0$ such that $P \cdot Y \geq 0$ for every $Y \in D$. Thus, $P \cdot (X - I(X)\mathbb{I}_S) \geq 0$ and $-P \cdot (X - I(X)\mathbb{I}_S) \geq 0$, implying that $P \cdot (X - I(X)\mathbb{I}_S) = 0$ for every $X \in \mathcal{Y}$. Therefore, $I(X) = P \cdot X$ for every $X \in \mathcal{Y}$. In particular, $P \cdot \mathbb{I}_S = 1$. Thus, P is a probability vector.

Fix a function X. Suppose that $Y_1 \leq Y_2 + X$, where Y_1 and Y_2 are SFaF. Then, by (1), $I(Y_1) \leq I(Y_2 + X)$. By (1) and (3), the right-hand side is equal to $I(Y_2) + I(X)$. This implies that $I(Y_1) - I(Y_2) \leq I(X)$. By Lemma 6, $\mathcal{Y}$ is convex and since the integral is homogenous, this implies that $\int X dP_{\mathcal{Y}} \leq I(X)$.

By (2) there exists $0 \leq \alpha < 1$ such that $[\alpha X + (1-\alpha)\mathbb{I}_S]$ is SFaF. Thus, $I([\alpha X + (1-\alpha)\mathbb{I}_S]) = I(\alpha X + (1-\alpha)\mathbb{I}_S) = I(\alpha X) + I((1-\alpha)\mathbb{I}_S)$ (the second equality is by (3)). This implies that $I([\alpha X + (1-\alpha)\mathbb{I}_S]) - I((1-\alpha)\mathbb{I}_S) = I(\alpha X)$. Since both $[\alpha X + (1-\alpha)\mathbb{I}_S]$ and $(1-\alpha)\mathbb{I}_S$ are SFaF, and $[\alpha X + (1-\alpha)\mathbb{I}_S] - (1-\alpha)\mathbb{I}_S \leq X$, we obtain (using (4)) that $\int X dP_{\mathcal{Y}} = I(X)$. ■

Proof of Proposition 2. The 'only if' direction is straightforward. As for the 'if' direction, note first that if E is FaF and if F does not intersect E, then $v(E) + v(F) = v(E \cup F)$. A set E is *minimal positive FaF* if $v(E) > 0$ and if $F \subsetneqq E$ implies $v(F) = 0$. Since S is finite, every set E contains at least one minimal positive FaF.

Secondly, let E and F be two minimal positive FaF's. Then, $E \cap F = \emptyset$. Indeed, $v(E \cup F) = v(E \cup (F \setminus E)) = v(E) + v(F \setminus E)$. If $E \cap F \neq \emptyset$, then $F \setminus E \subsetneqq F$ and therefore, $v(F \setminus E) = 0$. Thus, $v(E \cup F) = v(E)$. By a similar argument, $v(E \cup F) = v(F)$. Due to minimality, $E \cap F$ is a strict subset of both E and F. Thus, $v(E \cap F) = 0$ and therefore, $v(E) + v(F) = v(E \cup F) + 0$. This implies that $2v(E) = v(E)$, which contradicts the fact that $v(E) > 0$.

Let $\mathcal{A}' = \{A_1, ..., A_\ell\}$ be the set of all minimal positive FaF. The intersection of any two sets in $\mathcal{A}'$ is empty. It has been shown that v is additive over the sets in $\mathcal{A}'$. Denote by S' the union of $\mathcal{A}'$. It remains to show that $v(S) = v(S')$. Note that $v(S \setminus S') = 0$ since otherwise a minimal positive FaF would be included in $S \setminus S'$ (and therefore would be part of $\mathcal{A}'$). Since each of the A_i's is FaF, $v(A_1 \cup (S \setminus S')) = v(A_1) + v(S \setminus S') = v(A_1)$, then $v(A_2 \cup (A_1 \cup (S \setminus S'))) = v(A_2) + v(A_1)$ and so forth, until one obtains $v(S) = v(S')$, as desired.

Finally, let $\mathcal{A}$ be the sub-algebra generated by $\mathcal{A}' \cup \{S \setminus S'\}$. Now, define $P(A) = v(A)$ for every $A \in \mathcal{A}'$, extend P in a linear manner to $\mathcal{A}$, and for every $A \notin \mathcal{A}$ set $P(A) = \max_{B \subseteq A;\ B \in \mathcal{A}} P(B)$. It is clear that for every $A \subseteq S$, $P(A) \leq v(A)$, with equality for every $A \in \mathcal{A}$. To show a universal equality one can apply the sequential argument used in the previous paragraph. ■