Cognitive Foundations of Probability

By

Itzhak Gilboa and David Schmeidler

December 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1340

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Cognitive Foundations of Probability*

Itzhak Gilboa[†] and David Schmeidler[‡]

July 1999 – Revised June 2001

Abstract

Prediction is based on past cases. We assume that a predictor can rank eventualities according to their plausibility given any memory that consists of repetitions of past cases. In a companion paper, we show that under mild consistency requirements, these rankings can be represented by numerical functions, such that the function corresponding to each eventuality is linear in the number of case repetitions. In this paper we extend the analysis to rankings of events. Our main result is that a cancellation condition a la de Finetti implies that these functions are additive with respect to union of disjoint sets. If the set of past cases coincides with the set of possible eventualities, natural conditions are equivalent to ranking events by their empirical frequencies. More generally, our results may describe how individuals form probabilistic beliefs given cases that are only partially pertinent to the prediction problem at hand, and how this subjective measure of pertinence can be derived from likelihood rankings.

1 Introduction

The Bayesian approach holds that, facing uncertainty, one should form a prior and, given new information, update it according to Bayes rule. It relies on

*We wish to thank Didier Dubois, Roger Myerson, Peter Wakker, and Peyton Young for conversations and comments that motivated and influenced this work. We also thank an anonymous referee for helpful comments. This work was partially supported by ISF Grant No. 790/00.

[†]Tel-Aviv University and Cowles Foundation, Yale University. igilboa@post.tau.ac.il

[‡]Tel-Aviv University and The Ohio State University. schmeid@post.tau.ac.il

sound axiomatic foundations: Ramsey (1931), de Finetti (1937), and Savage (1954) argue that Bayesian expected utility maximization is the only normatively acceptable decision rule, and that in-principle-observable preferences can uniquely define a prior. Probabilities have also been axiomatically derived from qualitative plausibility judgments, where the latter are modeled as binary relations (see Kraft, Pratt, and Seidenberg (1959), Krantz, Luce, Suppes, and Tversky (1971), Fine (1973), Fishburn (1986)) or as propositions (see Fagin, Halpern, and Megiddo (1990), Fagin and Halpern (1994), Aumann (1995), Heifetz and Mongin (1999)).[1] But these axiomatic derivations do not explicitly model the information based on which a prior is formed. Further, they do not attempt to provide an account of this cognitive process. Thus, the axiomatizations of Bayesian beliefs and of Bayesian expected utility maximization may convince one that one would like to be Bayesian, but they do not provide the self-help tools that are needed to become a practising Bayesian.

The goal of this paper is to model explicitly the link between factual knowledge and derived beliefs. A special case of such a model should be a frequentist approach: when observing an experiment that is repeated under (seemingly) identical conditions, one may use past empirical frequencies as probabilities of future occurrences. But we wish to discuss also situations that are not repeated under precisely the same conditions. For instance, a physician encounters a new patient, who is not identical to any past patient she has treated, but may be similar to some. How is she to form a prior probability over the outcome of various treatments?

In a companion paper (Gilboa and Schmeidler (1999)) we suggest and axiomatize a rule for ranking of possible eventualities: each past case c and each possible eventuality x are assigned a number $v(x, c)$, and, given a collection of cases M, eventuality x is considered more likely than eventuality

[1]The idea of postulating qualitative probabilities as a basis for numeric probabilities originates with de Finetti, and consists a main step in Savage's derivation of subjective probabilities.

y if and only if

$$\sum_{c\in M} v(x,c) \geq \sum_{c\in M} v(y,c) \ ,$$

This rule can be viewed as a description of a cognitive process by which the predictor decides which of two eventualities is more likely. But it falls short of generating a prior over all possible events. It does not assume that the eventualities have any logical or algebraic structure. Further, the axiomatic derivation demands that, for every four eventualities, every permutation is a possible ranking (for an appropriately chosen M). This condition is counter-intuitive if eventualities are endowed with some additional structure. Specifically, if y is logically derived from x, one cannot expect x to be strictly more likely than y.

In this paper we deal with the case in which the objects of the likelihood rankings are events in a given algebra. We first show how, in this set-up, one may appropriately weaken the axioms of Gilboa and Schmeidler (1999) to obtain a result for the ranking of pairs events, none of which is included in the other. We then proceed to impose an additional condition of event cancellation a la de Finetti. Our main result is that this condition is equivalent to the condition that the functions $v(\cdot, c)$, for each case c, are additive with respect to the union of disjoint sets. It therefore describes the way that a predictor, who is committed to this cancellation condition, may form probabilistic beliefs over events given any possible memory.

We then proceed to test our model in the benchmark example of frequentism. That is, we assume that each case observed in the past can only be one of the possible eventualities in the problem at hand. Under this structural assumption it is natural to state two additional assumptions on plausibility rankings, which are shown to be equivalent to frequentism, namely, to ranking events by their empirical frequencies.

While it is reassuring to know that frequentism is a special case of our model, we consider it a conceptually simple problem. The more interesting problems, as in the example of the physician above, are those in which each past case is pertinent to the present problem to a certain subjective degree. Our results show how one may form probabilistic beliefs based on partially relevant information. Conversely, they also show how qualitative "at least as plausible as" comparisons may be used to elicit the subjective similarity judgments, and when these can be assumed additive with respect to set union.

The rest of this paper is organized as follows. Section 2 contains the main results. Sub-section 2.1 quotes relevant results from a companion paper. Sub-section 2.2 presents adapts the representation result to an algebra of events, while sub-section 2.3 contains the result about additivity of the set functions $v(\cdot, c)$. Section 3 deals with the situation in which the set of past cases coincides with the set of possible eventualities, and how frequentism then follows as a special case of our approach. Finally, an appendix contains all proofs.

2 Main Results

2.1 States

In this sub-section we describe a result that is proven in a companion paper, and which is the basis of the analysis that follows. Consider a prediction problem, in which one is asked to rank *eventualities* in a non-empty set X. For concreteness, one may think of these eventualities as states of the world throughout this sub-section. But the axioms that we use here do not presuppose that the eventualities to be ranked are mutually exclusive or exhaustive. Indeed, the result reported in this sub-section will later be used for various collections of events.

The predictor is equipped with knowledge of cases, facts, observations, or stories. Let M be a finite and non-empty set of *cases*, representing the predictor's knowledge. The person (or the machine) who is supposed to come up with predictions is assumed to have a well-defined "at least as likely as" relation on X, that presumably relies on M. Hence, for a different collection of cases the predictor may have a different "at least as likely as" relation. We assume that such a relation is given not only for the actual state of knowledge, but also for all hypothetical ones, that are generated from it by replication of cases.

Formally, consider the set of repetitions of cases $\mathbb{J} = \mathbb{Z}_+^M = \{I | I : M \to \mathbb{Z}_+\}$ where $\mathbb{Z}_+$ denotes the non-negative integers. For simplicity, we will refer to elements of $\mathbb{J}$ as *memories*. We assume that for every $I \in \mathbb{J}$ the predictor has a binary relation "at least as likely as" $\succsim_I$ on X (i.e., $\succsim_I \subseteq X \times X$).

Algebraic operations on $\mathbb{J}$ are performed pointwise. We define $\succ_I$ and $\approx_I$ to be the asymmetric and symmetric parts of $\succsim_I$, as usual.

We will use the following axioms:

A1 Order: For every $I \in \mathbb{J}$, $\succsim_I$ is complete and transitive on X.

A2 Combination: For every $I, J \in \mathbb{J}$ and every $x, y \in X$, if $x \succsim_I y$ ($x \succ_I y$) and $x \succsim_J y$, then $x \succsim_{I+J} y$ ($x \succ_{I+J} y$).

A3 Archimedeanity: For every $I, J \in \mathbb{J}$ and every $x, y \in X$, if $x \succ_I y$, then there exists $k \in N$ such that $x \succ_{kI+J} y$.

Observe that in the presence of A2, A3 also implies that for every $I, J \in \mathbb{J}$ and every $x, y \in X$, if $x \succ_I y$, then there exists $l \in \mathbb{N}$ such that for all $k \geq l$, $x \succ_{kI+J} y$.

Axiom 1 simply requires that, given any conceivable memory, the decision maker's preference relation over acts is a weak order. Axiom 2 states that if eventuality x is more plausible than eventuality y given two disjoint memories, x should also be more plausible than y given the combination of these memories. In our set-up, combination (or concatenation) of memories takes the form of adding the number of repetitions of each case in the two memories. Axiom 3 is a continuity, or an Archimedean axiom. It states that if, given the memory I, the predictor believes that eventuality x is strictly more plausible than y, then, no matter what her ranking is for another memory, J, there is a number of repetitions of I that is large enough to overwhelm the ranking induced by J.

We also need a diversity axiom that is not necessary for the functional form we would like to derive. While the theorem we present is an equivalence theorem, it characterizes a more restricted class of plausibility rankings than those discussed in the introduction. Specifically, we require that for any four eventualities, there is a memory that would distinguish among all four of them.

A4 Diversity: For every list (x, y, z, w) of distinct elements of X there exists $I \in \mathbb{J}$ such that $x \succ_I y \succ_I z \succ_I w$. If $|X| < 4$, then for any strict ordering of the elements of X there exists $I \in \mathbb{J}$ such that $\succ_I$is that ordering.

Finally, we need the following definition: a matrix of real numbers is called *diversified* if no row in it is dominated by an affine combination of three (or fewer) other rows in it. Formally:

Definition: A matrix $v : X \times Y \to \mathbb{R}$, where $|X| \geq 4$, is *diversified* if there are no distinct four elements $x, y, z, w \in X$ and $\lambda, \mu, \theta \in \mathbb{R}$ with $\lambda + \mu + \theta = 1$ such that $v(x, \cdot) \leq \lambda v(y, \cdot)) + \mu v(z, \cdot) + \theta v(w, \cdot)$. If $|X| < 4$, v is *diversified* if no row in v is dominated by an affine combination of the others.

We now quote a result of a previous work which will be used in this paper.

Theorem 2.1 *(Gilboa and Schmeidler (1999, 2001)): Let X, M, and $\{\succsim_I\}_{I \in \mathbb{J}}$ be given as above. Then the following two statements are equivalent:*

(i) $\{\succsim_I\}_{I \in \mathbb{J}}$ satisfy A1-A4;

(ii) There is a diversified matrix $v : X \times M \to \mathbb{R}$ such that:

$$(**) \quad \begin{cases} \text{for every } I \in \mathbb{J} \text{ and every } x, y \in X, \\ x \succsim_I y \quad \text{iff} \quad \sum_{c \in M} I(c) v(x, c) \geq \sum_{c \in M} I(c) v(y, c) \ , \end{cases}$$

***Furthermore**, in this case the matrix v is unique in the following sense: v and u both satisfy $(**)$ iff there are a scalar $\lambda > 0$ and a matrix $\beta : X \times M \to \mathbb{R}$ with identical rows (i.e., with constant columns) such that $u = \lambda v + \beta$.*

This theorem has several applications mentioned in Gilboa-Schmeidler (1999, revised version 2001). In particular, it can be viewed as axiomatizing kernel methods for estimation of density functions, as well as for classification problems. The theorem can also be interpreted as an axiomatization of maximum likelihood estimation: assume that X is a set of theories, or general rules one is to rank according to plausibility given memory. Axioms A1-4 appear reasonable for this case, and one can derive the representation $(**)$. If we shift the weights $v(x, c)$ so that they are all negative, they can be interpreted as logarithms of the conditional probability of case c given theory x. Thus Theorem 2.1 can be viewed as an axiomatization of ranking theories, or probability distributions, based on the likelihood function, together with a derivation of the conditional probabilities used in the likelihood function.

2.2 Events

The predictions discussed in sub-section 2.1 are abstract eventualities, lacking any logical or algebraic structure. It is natural to ask how similarity-based ranking of prediction relates to basic logical or set operations. In an attempt to address this question, we focus here on the case in which the alternatives to be ranked are events.

Let Ω be a state space. Let Σ be an algebra of events on Ω. Assume that Σ contains all singletons. Assume further that Ω contains at least 5 states. We assume that a "at least as likely as" relation between events in Σ that are not included in each other. More precisely, two events $A, B \in \Sigma$ are said to be *non-included* if $A \backslash B, B \backslash A \neq \emptyset$. We assume that only such pairs are ranked. In particular, we are interested only in proper non-empty subsets of Ω in Σ. For reasons that will be clarified in the proof, it is convenient to rule out of the discussion all proper subsets of Ω whose complement is a singleton. We therefore focus on $\Sigma' = \{A \in \Sigma \mid A \neq \emptyset \text{ and } |A^c| > 1\}$.

The application of Theorem 2.1 to the case of events is not immediate, because we only assume ranking between non-included events, and because the diversity axiom would require some modifications. We start by re-stating the first three axioms for the case at hand.

A1* Order: For every $I \in \mathbb{J}$, for every pair of non-included events $A, B \in \Sigma'$, $A \succsim_I B$ or $B \succsim_I A$. Further, if $A, B, C \in \Sigma'$ are pairwise non-included, then $A \succsim_I B$ and $B \succsim_I C$ imply $A \succsim_I C$.

A2* Combination: For every $I, J \in \mathbb{J}$ and every pair of non-included events $A, B \in \Sigma'$, if $A \succsim_I B$ $(A \succ_I B)$ and $A \succsim_J B$, then $A \succsim_{I+J} B (A \succ_{I+J} B)$.

A3* Archimedeanity: For every $I, J \in \mathbb{J}$ and every pair of non-included events $A, B \in \Sigma'$, if $A \succ_I B$, then there exists $k \in N$ such that $A \succ_{kI+J} B$.

Next we turn to the diversity axiom used in sub-section 2.1. Observe that, as stated, it cannot hold for any four-tuple of events. First, one cannot expect an event A to be strictly more plausible than an event B if $A \subseteq B$. Second, if one ranks plausibility according to a probability measure, and if $1_A + 1_B \leq 1_C + 1_D$ or $1_A + 1_C \leq 1_B + 1_D$ (where 1_E denotes the indicator function of $E \in \Sigma$), one cannot expect any memory to induce the ranking $A \succ_I B \succ_I C \succ_I D$. The proposition that follows shows that the exceptions above are the only ones.

Proposition 2.2 *Suppose that (A, B, C, D) are four events in a measurable space (Ω, Σ). Then there exists a probability measure P on Σ such that $P(A) > P(B) > P(C) > P(D)$ iff*

(i) No event in the list (A, B, C, D) is a subset of a follower in the list; and

(ii) Neither $1_A + 1_B \leq 1_C + 1_D$ nor $1_A + 1_C \leq 1_B + 1_D$.

It follows that, for a quadruple of pairwise non-including events, condition (ii) is necessary and sufficient for the existence of a measure that strictly ranks the four events in the given order. The proposition above motivates the following definitions: $(A, B, C, D) \in \Sigma'^4$ is a list of orderly differentiated events if no event in the list is a subset of a follower in the list, and neither $1_A + 1_B \leq 1_C + 1_D$ nor $1_A + 1_C \leq 1_B + 1_D$. The events $\{A, B, C, D\}$ are *properly differentiated* if every permutation thereof generates a list of orderly differentiated events. In order not to rule out rankings that agree with probability measures, we will restrict the requirement of diversity as follows:

A4* Restricted Diversity: For every list of orderly differentiated events (A, B, C, D), there exists $I \in \mathbb{J}$ such that $A \succ_I B \succ_I C \succ_I D$.

We can now state

Theorem 2.3 *Under the structural assumptions above, the following two statements are equivalent:*

(i) $\{\succsim_I\}_{I\in\mathbb{J}}$ *satisfy A1*-A4*;*

(ii) There is a diversified matrix $v : \Sigma' \times M \to \mathbb{R}$ *such that:*

$$(**) \quad \begin{cases} \text{for every } I \in \mathbb{J} \text{ and every pair of non-included events } A, B \in \Sigma', \\ A \succsim_I B \quad \text{iff} \quad \sum_{c\in M} I(c)v(A, c) \geq \sum_{c\in M} I(c)v(B, c), \end{cases}$$

Furthermore, in this case the matrix v *is unique in the following sense:* v *and* w *both satisfy* $(**)$ *iff there are a scalar* $\lambda > 0$ *and a matrix* $u : \Sigma' \times M \to \mathbb{R}$ *with identical rows (i.e., with constant columns) such that* $w = \lambda v + u$.

2.3 Additivity

Theorem 2.3 states that a method that ranks events by their likelihood, given any possible repetition of known cases, has to be equivalent to a numerical ranking where the number attached to each event is a linear function of the numbers of case repetitions. One naturally wonders, what would it take to make these numbers probabilities. That is, when is there a probability measure μ_c for each case c, such that memory I induces the same ranking of events as the measure $\sum_{c\in M} I(c)\mu_c$?

Obviously, a necessary condition for a representation by additive measures is that, for every $I \in \mathbb{J}$, $\succsim_I$ satisfies de Finetti's cancellation axiom: for every three events A, B, C such that $(A \cup B) \cap C = \emptyset$, we have $A \succsim_I B \Leftrightarrow A \cup C \succsim_I B \cup C$. A key result is that, if we impose this condition (restricted to non-included events) on top of the conditions of Theorem 2.3, the resulting matrix v can be normalized so that it is additive in events, namely, so that $v(A \cup B, \cdot) = v(A, \cdot) + v(B, \cdot)$ whenever $A \cap B = \emptyset$. Moreover, in this case we obtain uniqueness of the representation up to multiplication by a positive constant. Formally, we introduce the following axiom:

A5 Cancellation: For every $I \in \mathbb{J}$, and for every three pairwise non-included events A, B, C such that $(A \cup B) \cap C = \emptyset$, we have $A \succsim_I B \Leftrightarrow A \cup C \succsim_I B \cup C$.

Theorem 2.4 *Under the structural assumptions above, the following two statements are equivalent:*

(i) $\{\succsim_I\}_{I\in\mathbb{J}}$ *satisfy A1*-A4* and A5;*

(ii) There are finite, signed, and finitely additive measures $\{\mu_c\}_{c\in M}$ *such that:*

$$(**)\quad \begin{cases} \text{for every } I \in \mathbb{J} \text{ and every pair of non-included events } A, B \in \Sigma', \\ \qquad A \succsim_I B \quad \text{iff} \quad \sum_{c\in M} I(c)\mu_c(A) \geq \sum_{c\in M} I(c)\mu_c(B)\ , \end{cases}$$

and, for every list of orderly differentiated events (A, B, C, D)*, the vectors* $(\mu_c(A))_c$*,*$(\mu_c(B))_c$*,*$(\mu_c(C))_c$*, and* $(\mu_c(D))_c$ *define a diversified matrix.*

Furthermore, in this case the measures $\{\mu_c\}_{c\in M}$ *are unique up to multiplication by a positive number.*

Remark 2.5 Theorem 2.4 may not hold if Ω contains less than 5 states.

The statement (and proof) of Theorem 2.4 does not restrict Ω to be finite. Yet, the restricted diversity axiom may do so. For instance, it is easy to see that, if the measures μ_c are non-negative, A4* can be satisfied only if Ω is, indeed, finite. However, one may have versions of the theorem that allow infinite Ω (say, with an infinite set of cases).

Theorem 2.4 only guarantees representation by signed measures. Indeed, since we only use comparisons of pairwise non-included events, the data $\{\succsim_I\}_{I\in\mathbb{J}}$ do not imply that likelihood rankings are monotone with respect to set inclusion. One may require that, for each $I \in \mathbb{J}$, $\succsim_I$, be a *qualitative probability* according to de Finetti (1931), namely that:

(i) $\succsim_I$ is complete and transitive on Σ;

(ii) for every three events A, B, C such that $(A \cup B) \cap C = \emptyset$, we have $A \succsim_I B \Leftrightarrow A \cup C \succsim_I B \cup C$;

(iii) for every event A, $A \succsim_I \emptyset$;

(iv) $\Omega \succ_I \emptyset$.

This condition strengthens both A1* and A5. One may conjecture that imposing it would yield a representation such as in $(**)$ of Theorem 2.4 for *all* pairs of events. As stated, the answer cannot be in the affirmative since any numerical representation by $\{\mu_c\}_{c\in M}$ would yield $A \approx_0 B$ for all $A, B \in \Sigma$ (where $0 \in \mathbb{J}$ denotes the memory in which all cases appear zero times), contradicting (iv). A more natural condition to impose is, therefore,

A1' Qualitative Probability: For every $I \in \mathbb{J}$, $\succsim_I$ satisfies (i)-(iii) of the definition above, and if $I \neq 0$, $\succsim_I$ also satisfies (iv).

Yet, even with this weakening the conjecture is false:

Remark 2.6 Assume that $\{\succsim_I\}_{I\in\mathbb{J}}$ satisfy A1', A2*, A3*, A4*. It is possible that the signed measures $\{\mu_c\}_{c\in M}$ obtained in Theorem 2.4 fail to be non-negative.

3 Frequentism

The framework of Section 2 does not assume any formal relationship between past cases and states of the world. Indeed, one of the strengths of the approach outlined above is that any such relationships may be inferred from plausibility rankings given various memories, rather than assumed a-priori. Still, an interesting special case, which is also an important test case, is the situation where memory consists only of past occurrences of the same states that are now possible. For instance, one may be asked to rank the possible outcomes of a roll of a die based on empirical frequencies of these outcomes in past rolls of the same die. It would be reassuring to know that our approach is compatible with frequentism, i.e., that the numerical rankings derived in sub-section 2.3 may boil down to relative empirical frequencies in this case.

Assume, then, that $M = \Omega = \{1, \ldots, n\}$, where $I \in \mathbb{J}$ is interpreted as the empirical frequencies of the possible outcomes. Assume that the relations in $\{\succsim_I\}_{I\in\mathbb{J}}$ are qualitative probability relations. We impose two additional assumptions. The first is a symmetry axiom, stating that the names of the outcomes are immaterial. The second is specificity axiom, requiring that an outcome that has never been observed does not increase the plausibility of events containing it. For the symmetry axiom, we introduce the following notation: let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation. For $I \in \mathbb{J}$, define $I \circ \pi \in \mathbb{J}$ by $I \circ \pi(c) = I(\pi(c))$ and for $A \in \Sigma$ let $A_\pi \in \Sigma$ be defined by $A_\pi = \{\pi^{-1}(i) \mid i \in A\}$. Using this notation,

A6 Symmetry: For every permutation π, every $I \in \mathbb{J}$, and every $A, B \in \Sigma$, $A \succsim_{I\circ\pi} B$, iff $A_\pi \succsim_I B_\pi$.

Next we introduce

A7 Specificity: Assume that for $j \in \Omega$ and $I \in \mathbb{J}$, $I(j) = 0$. For every $A, B \in \Sigma$, such that $B \succsim_I A$, we also have $B \succsim_I A \cup \{j\}$.

Theorem 3.1 *Assume that $n \geq 5$ and that $\{\succsim_I\}_{I \in \mathbb{J}}$ satisfy A1′, A2*, A3*, A4*,A6, and A7. Let $\{\mu_i\}$ be the measures provided by Theorem 2.4. Then $(\mu_i(\{j\}))_{1 \leq i,j \leq n}$ is the identity matrix, up to multiplication by a positive number.*

When $n < 5$ Theorem 2.4 does not provide a representation of $\{\succsim_I\}_{I \in \mathbb{J}}$ by measures μ_i. Yet, we can simply assume a representation along the lines of Theorem 2.4 and obtain a similar result:

Remark 3.2 Let $\{ \succsim_I\}_{I \in \mathbb{J}}$ be a collection of binary relations on $\Sigma = 2^\Omega$, and let $\{\mu_i\}$ be non-negative measures on Ω such that for every $I \in \mathbb{J}$ and every $A, B \in \Sigma$,

$$A \succsim_I B \quad \text{iff} \quad \sum_{c \in M} I(c)\mu_c(A) \geq \sum_{c \in M} I(c)\mu_c(B) \quad .$$

Assume that $\{ \succsim_I\}_{I \in \mathbb{J}}$ satisfy A1′,A6 and A7. Then $(\mu_i(\{j\}))_{1 \leq i,j \leq n}$ is the identity matrix, up to multiplication by a positive number.

Observe that A6 and A7 are obviously necessary for the two results.

Appendix: Proofs and Related Analysis

Proof of Proposition 2.2:

Let there be given four events A, B, C, D in a measurable space (Ω, Σ). It suffices to consider the minimal algebra containing these events, which is finite. Assume that the atoms in this algebra are $\Omega' = \{1, \ldots, n\}$. It is easy to see that there exists a probability measure P on Σ such that $P(A) > P(B) > P(C) > P(D)$ iff the following LP problem is feasible:

$$
\begin{array}{lll}
(P) & Min_{x \in \mathbb{R}^M_+} 0 \cdot x & \\
& s.t. & (1_A - 1_B) \cdot x \geq 1 \\
& & (1_B - 1_C) \cdot x \geq 1 \\
& & (1_C - 1_D) \cdot x \geq 1 \ .
\end{array}
$$

Problem (P) is feasible iff its dual is bounded. The dual is

$$
\begin{array}{llll}
(D) & Max & \alpha + \ \beta + \ \gamma & \\
& s.t. & \alpha(1_A - 1_B) + \ \beta(1_B - 1_C) + \ \gamma(1_C - 1_D) \leq 0 & (*) \\
& & \alpha, \beta, \gamma \geq 0 &
\end{array}
$$

Clearly, (D) is bounded iff it is bounded by zero, and this is the case iff its only feasible point is $\alpha, \beta, \gamma = 0$.

We now prove that conditions (i) and (ii) imply these equalities. Assume, by way of negation, that $(*)$ holds and let $\alpha, \beta, \gamma \geq 0$ but not all three are equal to zero.

Observe first that if exactly one of α, β,and γ is positive, it follows that at least one of the inclusions; $A \subseteq B$, $B \subseteq C$, or $C \subseteq D$ holds, in contradiction to (i).

Next assume that exactly two of α, β, and γ are positive. Suppose first that $\beta = 0$. Condition (ii) implies that for some k , $1_A(k) + 1_C(k) > 1_B(k) + 1_D(k)$. Since the indicator vectors take only the values 1 or 0, there are five possibilities in which this inequality may be satisfied: $1+1>1+0$, $0+1$, $0+0$ or $1+0$, $0+1>0+0$. Evaluating $(*)$ at k for these five possibilities leads to one of the following: $\alpha-\alpha+\gamma\leq 0$, $\alpha+\gamma-\gamma\leq 0$, $\alpha+\gamma\leq 0$, $\alpha\leq 0$, or $\gamma\leq 0$. All these inequalities are inconsistent with positivity of α and γ , so this case is ruled out. If $\gamma=0$, consider $i\in A\backslash C$ (whose existence follows from (i)). Now $(*)$ implies $\alpha\leq 0$ or $\alpha-\beta+\beta\leq 0$. Both contradict positivity of α. The last case is $\alpha=0$. Let $l\in B\backslash D$ (again, such a state exists by (i)). In this case $(*)$ implies $\beta\leq 0$ or $\beta-\beta+\gamma\leq 0$ – a contradiction to $\beta,\gamma>0$.

So we are left with the case where all three, α, β, and γ , are positive. Since the first inequality in (ii) does not hold, there exists i for which $1_A(i)+1_B(i)>1_C(i)+1_D(i)$. We consider the five possibilities, as above: $1+1>1+0$, $0+1$, $0+0$ or $1+0$, $0+1>0+0$. Substituting the corresponding values in $(*)$ we get that at least one of the following five inequalities holds: $\alpha-\alpha+\beta-\beta+\gamma\leq 0$, $\alpha-\alpha+\beta-\gamma\leq 0$, $\alpha-\alpha+\beta\leq 0$, $\alpha\leq 0$, $-\alpha+\beta\leq 0$. The positivity of α , β, and γ leaves only $\beta\leq\alpha$ or $\beta\leq\gamma$.

By (i), there is a state $j\in A\backslash D$. Again, using $(*)$, we get one of the following four inequalities: $\alpha\leq 0$, $\alpha-\alpha+\beta\leq 0$, $\alpha-\beta+\gamma\leq 0$, and $\alpha-\alpha+\beta-\beta+\gamma\leq 0$, depending on whether j belongs to B and to C or not. Three of the inequalities are directly inconsistent with positivity of α, β, and γ. The fourth, namely, $\alpha+\gamma\leq\beta$, is also inconsistent with positivity when coupled with either of $\beta\leq\alpha$ or $\beta\leq\gamma$ obtained previously. This concludes the proof that (i) and (ii) suffice for the existence of a probability measure P such that $P(A)>P(B)>P(C)>P(D)$. The necessity of (i) and (ii) is obvious. □□

Proof of Theorem 2.3:

We will construct the numerical representation by "patching" together numerical representations for subsets of events that are properly differentiated. In doing so, a few auxiliary results will be of help. We start by the following definition. Suppose that for a subset of events $\Delta\subseteq\Sigma'$ there is a matrix $v^{\Delta} : \Delta \times M \to \mathbb{R}$. Let Δ' be a subset of Δ. We say that v^{Δ} *ranks* Δ' if for every non-included $E, F \in \Delta'$, and every $I \in \mathbb{J}$,

$$E \succsim_I F \quad \text{iff} \quad \sum_{c \in M} I(c) v^{\Delta}(E, c) \geq \sum_{c \in M} I(c) v^{\Delta}(F, c) \ .$$

In order to extend a numerical representation v^{Δ} to a larger set Δ, we would like to know that such a representation is unique on relatively small subsets Δ'. For instance, when we consider triples of pairwise non-included events A, B, C, it would be nice to know that a function v that ranks $\{A, B, C\}$ is unique as in Theorem 2.1. For this one would need to have a diversity axiom for triples of events, namely that for any permutation thereof there exists an $I \in \mathbb{J}$ such that $\succ_I$ agrees with the given permutation. One would expect this to follow from the seemingly more powerful diversity assumption A4*, stated for all quadruples of orderly differentiated events. However, not every triple of pairwise non-included events can be complemented to a quadruple of orderly differentiated events. Consider, for instance, $n = 5$, $A = \{1, 2, 3\}$, $B = \{4\}$, $C = \{5\}$. These are pairwise non-included, but there is no event D that is pairwise non-included with respect to all of them.

This case is anomalous enough to deserve a definition. We say that three events $A, B, C \in \Sigma'$ *form an all-but-two partition* if two of them are singletons and the third is the complement of the (union of the) first two. We now state

Lemma 1: *Let $A, B, C \in \Sigma'$ be three pairwise non-included events that do not form an all-but-two partition. Then there exists $I \in J$ such that $A \succ_I B \succ_I C$.*

Proof: First observe that, since A, B, C are pairwise non-included, there exist probability measures P on Ω such that $P(A) > P(B) > P(C)$. We will shortly prove that there exists an event D that is non-included with respect to each of A, B, and C. We can choose a probability P such that $P(A) > P(B) > P(C)$ and that $P(D)$ differs from each of $\{P(A), P(B), P(C)\}$. This would mean, by Proposition 2.2, that one of the four lists, $\{(D,A,B,C),$ $(A,D,B,C),$ $(A,B,D,C),$ $(A,B,C,D)\}$ is orderly differentiated. We can then use the diversity axiom for that list to deduce the desired result.

We therefore wish to prove that there exists an event D that is non-included with respect to each of $A,B,$ and C. If there exists a state $i \in (A \cup B \cup C)^c$, then choosing $D = \{i\}$ will do. Assume, then, that $A \cup B \cup C = \Omega$. Next, if $A, B,$ and C are pairwise disjoint, since they do not form an all-but-two partition, it has to be the case that at least two of them contain more than one element. Assume, without loss of generality, that $\{i,j\} \subseteq A$ and that $\{k,l\} \subseteq B$. In this case, $D = \{i,k\}$ is non-included with respect to each of $A, B,$ and C.

We now deal with the cases where $A \cup B \cup C = \Omega$ and the three events are not pairwise disjoint. Assume that one of them is disjoint from the other two, say, $A \cap (B \cup C) = \emptyset$. Let $i \in B \backslash C$ and $j \in C \backslash B$. Since not all three events are disjoint, $B \cap C \neq \emptyset$, and it follows that $D = \{i,j\}$ is non-included with respect to all three.

We therefore assume that $A \cup B \cup C = \Omega$ and that each event intersects the union of the other two. Observe that this implies that none of A, B, C is a singleton. Assume that one of them is not contained in the union of the other two, say, $A \backslash (B \cup C) \neq \emptyset$. Then there exists $i \in A \backslash (B \cup C)$. Choose $j \in B \backslash A$ and let $D = \{i,j\}$.

Finally, we are left with the case where $A \cup B \cup C = \Omega$ and where each event is contained in the union of the other two. Hence every state in Ω is included in at least two of $\{A,B,C\}$. In this case $A^c = (B \cap C) \backslash A$; $B^c = (A \cap C) \backslash B$; and $C^c = (A \cap B) \backslash C$. Since A, B, C are in Σ', each of these pairwise disjoint events includes at least two elements. Construct D by selecting one of each. □

Lemma 2: *Let there be given two subsets $\sigma, \tau \subseteq \Sigma'$ with corresponding matrices $v^\sigma : \sigma \times M \to R$ and $v^\tau : \tau \times M \to R$. Assume that there are three pairwise non-included events $A, B, C \in \sigma \cap \tau$ that do not form an all-but-*

two partition, and that both v^σ and v^τ rank $\{A, B, C\}$. If, for all $c \in M$, $v^\sigma(A, c) = v^\tau(A, c)$ and $v^\sigma(B, c) = v^\tau(B, c)$, then also $v^\sigma(C, c) = v^\tau(C, c)$ for all $c \in M$.

Proof: In view of the previous lemma, the triple $\{A, B, C\}$ satisfies the conditions of Theorem 2.1. The conclusion follows from uniqueness result of the Theorem. $\square$

Lemma 3: *Let there be given a subset of events $\Delta \subseteq \Sigma'$ and a matrix $v^\Delta : \Delta \times M \to R$. Assume that $A, B, C, D \in \Delta$ are properly differentiated. If v^Δ ranks $\{A, C, D\}$ and $\{B, C, D\}$, then it also ranks $\{A, B\}$.*

Proof: Since A, B, C, D are properly differentiated, we can apply Theorem 2.1 to $\sigma \equiv \{A, B, C, D\}$, and conclude that there exists a matrix v^σ that ranks σ. Without loss of generality we may assume that $v^\sigma(C, c) = v^\Delta(C, c)$ and $v^\sigma(D, c) = v^\Delta(D, c)$ for all $c \in M$. Since A, B, C, D are properly differentiated, we know that none of $\{A, C, D\}$, $\{B, C, D\}$ forms an all-but-two partition. The previous remark therefore states that for all $c \in M$, $v^\sigma(A, c) = v^\Delta(A, c)$ and $v^\sigma(B, c) = v^\Delta(B, c)$ also hold. Since v^σ ranks $\{A, B\}$, so does v^Δ. $\square$

We now turn to the construction of a matrix $v^{\Sigma'}$ that ranks Σ'. The strategy is as follows: we start by constructing a representation for all pairs. (Recall that Σ includes all singletons, and therefore also all pairs.) We then extend it to singletons. Next we show that it can be extended to all events in Σ'.

Define $\Delta_2 = \{A \in \Sigma' | \, |A| = 2\}$.

Lemma 4: *There exists $v^{\Delta_2} : \Delta_2 \times M \to R$ that ranks Δ_2.*

Proof: Choose an element of Ω, and call it 1. We first consider $\Delta_2^1 = \{\{1, i\} | i \neq 1\}$. Any four events in Δ_2^1 are properly differentiated, and Theorem 2.1 can be applied to obtain a representation $v^{\Delta_2^1} : \Delta_2^1 \times M \to \mathbb{R}$ that ranks Δ_2^1.

Next we wish to extend $v^{\Delta_2^1}$ to v^{Δ_2} on all of Δ_2. Consider $A = \{i, j\}$ where $i, j \neq 1$. Any four events in $\Delta_2^1 \cup \{A\}$ are properly differentiated.

Hence Theorem 2.1 offers a unique definition of $v^{\Delta_2}(A,c)$ for $c \in M$.

We claim that v^{Δ_2} ranks Δ_2. Let there be given $A, B \in \Delta_2$. Notice that if at least one of them is in Δ_2^1, we already know that Δ_2^1 ranks $\{A, B\}$. Assume, then, that $A, B \in \Delta_2 \setminus \Delta_2^1$. Distinguish between two cases: (i) $A \cap B = \emptyset$; and (ii) $A \cap B \neq \emptyset$.

In Case (i) we have $A = \{i,j\}$, $B = \{k,l\}$ where i,j,k,l are distinct and differ from 1. Observe that any distinct four events out of $\Delta_2^1 \cup \{A, B\}$ are properly differentiated. It follows that there exists a matrix defined on $(\Delta_2^1 \cup \{A, B\}) \times M$ that ranks $\Delta_2^1 \cup \{A, B\}$, and that it coincides with our definition of v^{Δ_2}.

In Case (ii) we have $A = \{i,j\}$, $B = \{i,k\}$ where i,j,k are distinct and differ from 1. Observe that $\{\{1,i\},\{1,j\},\{i,j\},\{i,k\}\}$ are properly differentiated. By the Lemma 3, since v^{Δ_2} ranks $\{\{1,i\},\{1,j\},\{i,j\}\}$ and $\{\{1,i\},\{1,j\},\{i,k\}\}$, it also ranks $\{\{i,j\},\{i,k\}\}$. $\square$

Our next step is to extend v^{Δ_2} to singletons. Let $\overline{\Delta}_2 = \{A \in \Sigma' |\ |A| \leq 2\}$.

Lemma 5: *There exists $v^{\overline{\Delta}_2} : \overline{\Delta}_2 \times M \to R$ that ranks $\overline{\Delta}_2$.*

Proof: Let $v^{\overline{\Delta}_2}$ equal v^{Δ_2} on all pairs. We now extend it to all singletons, and then show that this extension indeed ranks $\overline{\Delta}_2$. Let there be given $i \in \Omega$. Choose distinct $j,k,l \neq i$. There is a unique definition of $v^{\overline{\Delta}_2}(\{i\},c)$ (for $c \in M$) such that $v^{\overline{\Delta}_2}$ ranks $\{\{i\},\{j,k\},\{j,l\}\}$. We first claim that $v^{\overline{\Delta}_2}$ thus defined ranks $\Delta_2 \cup \{\{i\}\}$. Indeed, for any event $B \in \Delta_2$ that differs from $\{j,k\},\{j,l\}$, $\{B,\{j,k\},\{j,l\}\}$ are pairwise non-included, and they do not form an all-but-two partition. Hence $v^{\overline{\Delta}_2}$ ranks $\{B,\{j,k\},\{j,l\}\}$. Further, if $i \notin B$, then $\{B,\{i\},\{j,k\},\{j,l\}\}$ are also properly differentiated, and, by Lemma 3, $v^{\overline{\Delta}_2}$ also ranks $\{B,\{i\}\}$.

Let this be the definition of $v^{\overline{\Delta}_2}(\{i\},c)$ for each $i \in \Omega$. We need to show that for every distinct $i,j \in \Omega$, $v^{\overline{\Delta}_2}$ ranks $\{\{j\},\{i\}\}$. Since $|\Omega| \geq 5$, there are two distinct $C, D \in \Delta_2$ that are disjoint from $\{i,j\}$. Thus, $\{\{i\},\{j\},C,D\}$ are properly differentiated, while $v^{\overline{\Delta}_2}$ ranks both $\{\{i\},C,D\}\}$ and $\{\{j\},C,D\}\}$, which completes the proof. $\square$

The following combinatorial lemma will prove useful.

Lemma 6: *Let A and B be two non-included events in Σ'. Then there are C and D in Σ', with $|C|, |D| = 2$, such that $\{A, B, C, D\}$ are properly differentiated.*

Proof: Assume without loss of generality that $|A| \geq |B|$, and distinguish between two cases: Case 1: $|A \backslash B| \geq 2$; Case 2: $|A \backslash B| = |B \backslash A| = 1$. In Case 1 set $i, j \in A \backslash B$, $i \neq j$, and $k \in B \backslash A$. Define $C = \{i, k\}$ and $D = \{j, k\}$. By direct verification one can check that the conclusion of the lemma holds.

Next consider Case 2. Since $|\Omega \backslash A|, |\Omega \backslash B| \geq 2$, there is a state $p \in \Omega \backslash (A \cup B)$. Assume first that there also is a state $q \neq p$, such that $q \in \Omega \backslash (A \cup B)$. Let $i \in A \backslash B$, and $k \in B \backslash A$. If A and B are singletons, there exists $j \in \Omega \backslash \{i, k, p, q\}$ and we can choose $C = \{j, p\}$ and $D = \{j, q\}$. Otherwise (namely, $A \cap B \neq \emptyset$) defining $C = \{i, p\}$ and $D = \{k, q\}$ results in the desired conclusion.

We are now left with Case 2 under the additional restriction that $|\Omega \backslash (A \cup B)| = 1$. Since $|\Omega| \geq 5$, we know that $|A \cap B| \geq 2$. Define $C = \{p, k\}$ and $D = \{p, l\}$, where $k \neq l$, $k, l \in A \cap B$. Once again, direct verification completes the proof. $\square$

Completion of the Proof of Theorem 2.3: We now proceed to define $v = v^{\Sigma'} : \Sigma' \times M \to \mathbb{R}$ that ranks Σ', as an extension of $v^{\overline{\Delta}_2}$. Let there be given an event $A \in \Sigma'$ with $|A| > 2$. Let i be an element in Ω that is not included in A, and let j, k be two elements that are in A. Since $\{A, \{i, j\}, \{i, k\}\}$ are pairwise non-included and they do not form an all-but-two partition, Theorem 2.1 applies to them and offers a unique definition of $v(A, c)$ (for all $c \in M$) such that v ranks $\{A, \{i, j\}, \{i, k\}\}$.

We now wish to show that v thus defined ranks $\{A, B\}$ for all non-included $A, B \in \Sigma'$. If $|A|, |B| \leq 2$, the result follows from the definition of v as an extension of $v^{\overline{\Delta}_2}$. Assume, then, that $|A| > 2$. We split the proof into three parts according to the number of elements in B.

First assume that $|B| = 2$. Recall that i is an element in Ω that is not included in A, and that j, k are two elements that are in A. There are four cases to check, according to whether $i \in B$ and whether A and B are disjoint. In all cases, direct verification shows that $\{A, B, \{i, j\}, \{i, k\}\}$ are properly differentiated, and Lemma 3 implies that v ranks $\{A, B\}$.

Next assume that $|B| = 1$, i.e., that $B = \{l\}$ where $l \notin A$. Choose $s \notin A \cup \{l\}$. Since j, k are in A, $\{A, B, \{s, j\}, \{s, k\}\}$ are properly differentiated, and Lemma 3 implies that v ranks $\{A, \{l\}\}$.

Finally, assume $|B| > 2$. By Lemma 6 there are C and D in Σ', with $|C|, |D| = 2$, such that $\{A, B, C, D\}$ are properly differentiated. We already know v ranks $\{A, C, D\}$ and $\{B, C, D\}$, since C and D are pairs. Again, Lemma 3 implies that v ranks $\{A, B\}$.

It is easy to see that, if for some list of orderly differentiated events (A, B, C, D), the vectors $(\mu_c(A)), (\mu_c(B))_c, (\mu_c(C))_c$, and $(\mu_c(D))_c$ form a matrix that is not diversified, then A4* is violated. This was also proven in detail in Gilboa-Schmeidler (1999). $\square$

The proof of sufficiency and of uniqueness are as in Gilboa-Schmeidler (1999). $\square\square$

Proof of Theorem 2.4:

The fact that (ii) implies (i) is immediate. We will show that (i) implies (ii) and the uniqueness result for the case of a finite algebra.

Assume, then, that $\Omega = \{1, \ldots, n\}$ (recall that $n \geq 5$) and that $\Sigma = 2^{\Omega}$, and let $\hat{v}$ be the matrix provided by Theorem 2.3. Set $w(\cdot) = \hat{v}(\{1, 2\}, \cdot) - \hat{v}(\{1\}, \cdot) - \hat{v}(\{2\}, \cdot)$, and define a matrix v by $v(A, \cdot) = \hat{v}(A, \cdot) + w(\cdot)$ so that $v(\{1, 2\}, \cdot) = v(\{1\}, \cdot) + v(\{2\}, \cdot)$. We wish to show that for this v, $v(A, \cdot) = \sum_{i \in A} v(\{i\}, \cdot)$ for every $A \in \Sigma'$. Observe that if we find such a v, it is unique up to positive multiplication, since a shift by a vector w can preserve additivity only if $w = 0$.

Some notation may prove useful. We will use event superscripts to denote rows in the matrix. Thus, v^A denotes the vector $v(A, \cdot)$. Also, for $A, B \in \Sigma'$, define $v^{A,B} = v^A - v^B$. Note that for any three pairwise non-included events

A, B, C, we have the *Jacobi identity* $v^{A,C} = v^{A,B} + v^{B,C}$. A key observation is the following

Lemma 1: *For every three pairwise non-included events A, B, C such that $(A \cup B) \cap C = \emptyset$, there exists a unique $\lambda > 0$ such that $v^{A\cup C, B\cup C} = \lambda v^{A,B}$.*

Proof: By the cancellation axiom we know that the set of memories I for which $A \cup C \succsim_I B \cup C$ is precisely the same set for which $A \succsim_I B$. The conclusion follows from the uniqueness result of Theorem 2.1 applied to the events A, B. $\square$

Next we show that, when we focus on a singleton $C = \{i\}$, the coefficient λ does not depend on the sets A, B:

Lemma 2: *For every $i \in \Omega$ there exists a unique $\lambda_i > 0$ such that, for every non-included $A, B \in \Sigma'$ such that $i \notin A \cup B$ and $|A|, |B| < n - 2$, $v^{A\cup\{i\}, B\cup\{i\}} = \lambda_i v^{A,B}$.*

Proof: Consider three events A, B, D that are pairwise non-included, none of which includes i, and none of which has more than $n - 3$ elements. We know that

$$v^{A\cup\{i\}, B\cup\{i\}} = v^{A\cup\{i\}, D\cup\{i\}} + v^{D\cup\{i\}, B\cup\{i\}}$$

applying Lemma 1 to each of the three elements above we obtain numbers $\lambda, \mu, \eta > 0$ such that

$$\begin{aligned} v^{A\cup\{i\}, B\cup\{i\}} &= \lambda v^{A,B} \\ v^{A\cup\{i\}, D\cup\{i\}} &= \mu v^{A,D} \\ v^{D\cup\{i\}, B\cup\{i\}} &= \eta v^{D,B} \end{aligned}$$

hence

$$\lambda v^{A,B} = \mu v^{A,D} + \eta v^{D,B}$$

or

$$v^{A,B} = \frac{\mu}{\lambda} v^{A,D} + \frac{\eta}{\lambda} v^{D,B} \ .$$

But since we also have the Jacobi identity

$$v^{A,B} = v^{A,D} + v^{D,B}$$

the restricted diversity axiom implies that $\lambda = \mu = \eta$.

Applying this result for the case that A, B, D are singletons, we conclude that there exists $\lambda_i > 0$ such that the conclusion holds for all singletons A, B. Next, for every $A \in \Sigma'$ such that $i \notin A$ and $|A| < n-2$, and every $j \notin A \cup \{i\}$, we also get $v^{A\cup\{i\},\{i,j\}} = \lambda_i v^{A,\{j\}}$. Similarly, consider such a set A, and choose $j, k \notin A \cup \{i\}$ and $l \in A$. Define $B = \{l, k\}$ and $D = \{j\}$. We obtain that $v^{A\cup\{i\},\{i,k,l\}} = \lambda_i v^{A,\{k,l\}}$. Finally, consider two non-included events $A, B \in \Sigma'$ such that $i \notin A \cup B$ and $|A|, |B| < n-2$. Choose $l \in A \backslash B$ and $k \in B \backslash A$. Define $D = \{l, k\}$ and apply the result above to these three events. The desired result follows. $\square$

Lemma 3: *The exists a unique $\lambda > 0$ such that $\lambda = \lambda_i$ for all $i \in \Omega$ (where λ_i is the coefficient defined by Lemma 2). Further, this λ satisfies, for every distinct $i, j, k, l \in \Omega$,*

$$v^{\{i,j\},\{k,l\}} = \lambda v^{\{i\},\{k\}} + \lambda v^{\{j\}\{l\}} \ .$$

Proof: By the Jacobi identity and Lemma 2,

$$v^{\{i,j\},\{k,l\}} = v^{\{i,j\},\{k,j\}} + v^{\{k,j\},\{k,l\}} = \lambda_j v^{\{i\},\{k\}} + \lambda_k v^{\{j\},\{l\}}.$$

Similarly,

$$v^{\{i,j\},\{k,l\}} = v^{\{i,j\},\{l,j\}} + v^{\{l,j\},\{k,l\}} = \lambda_j v^{\{i\},\{l\}} + \lambda_l v^{\{j\},\{k\}}$$

hence

$$\lambda_j v^{\{i\},\{k\}} + \lambda_k v^{\{j\},\{l\}} = \lambda_j v^{\{i\},\{l\}} + \lambda_l v^{\{j\},\{k\}}$$

or

$$\lambda_j v^{\{i\},\{k\}} = \lambda_j v^{\{i\},\{l\}} + \lambda_k v^{\{l\},\{j\}} + \lambda_l v^{\{j\},\{k\}} \ .$$

That is,

$$v^{\{i\},\{k\}} = v^{\{i\},\{l\}} + \frac{\lambda_k}{\lambda_j} v^{\{l\},\{j\}} + \frac{\lambda_l}{\lambda_j} v^{\{j\},\{k\}} \ .$$

But the Jacobi identity also implies

$$v^{\{i\},\{k\}} = v^{\{i\},\{l\}} + v^{\{l\},\{j\}} + v^{\{j\},\{k\}}$$

and, coupled with the diversity axiom this means that $\lambda_j = \lambda_k = \lambda_l$. $\square$

Until the end of the proof we reserve the symbol λ to the coefficient defined by Lemma 3.

Lemma 4: *For every distinct* $i \in \Omega \backslash \{1, 2\}$*,*

$$v^{\{1,i\}} = v^{\{1\}} + (1-\lambda) v^{\{2\}} + \lambda v^{\{i\}}$$
$$v^{\{2,i\}} = (1-\lambda) v^{\{1\}} + v^{\{2\}} + \lambda v^{\{i\}} \ .$$

Proof: By symmetry between 1 and 2, it suffices to prove the first equation. Since $v^{A,B} = v^A - v^B$ (applied to $A = \{1, i\}$ and $B = \{1, 2\}$) we get

$$v^{\{1,i\}} = v^{\{1,i\},\{1,2\}} + v^{\{1,2\}}.$$

By Lemma 3

$$= \lambda v^{\{i\},\{2\}} + v^{\{1,2\}}$$

and, using the fact $v^{\{1,2\}} = v^{\{1\}} + v^{\{2\}}$ and $v^{A,B} = v^A - v^B$ (applied to $A = \{i\}$ and $B = \{2\}$)

$$\begin{aligned} &= \lambda v^{\{i\}} - \lambda v^{\{2\}} + v^{\{1\}} + v^{\{2\}} \\ &= v^{\{1\}} + (1-\lambda) v^{\{2\}} + \lambda v^{\{i\}} \ . \end{aligned}$$

$\square$

Lemma 5: *For every distinct* $i, j \in \Omega\backslash\{1,2\}$,

$$v^{\{i,j\}} = (1-\lambda)v^{\{1\}} + (1-\lambda)v^{\{2\}} + \lambda v^{\{i\}} + \lambda v^{\{j\}} \ .$$

Proof: Since $v^{A,B} = v^A - v^B$ (applied to $A = \{i,j\}$ and $B = \{1,2\}$) we get

$$v^{\{i,j\}} = v^{\{i,j\},\{1,2\}} + v^{\{1,2\}}.$$

Using the Jacobi identity

$$v^{\{i,j\}} = v^{\{i,j\},\{1,i\}} + v^{\{1,i\},\{1,2\}} + v^{\{1,2\}}.$$

By Lemma 3 the right hand side equals

$$\lambda v^{\{j\},\{1\}} + \lambda v^{\{i\},\{2\}} + v^{\{1,2\}}.$$

Using the facts $v^{\{1,2\}} = v^{\{1\}} + v^{\{2\}}$ and $v^{A,B} = v^A - v^B$ we get

$$\begin{aligned} v^{\{i,j\}} &= \lambda v^{\{j\}} - \lambda v^{\{1\}} + \lambda v^{\{i\}} - \lambda v^{\{2\}} + v^{\{1\}} + v^{\{2\}} \\ &= (1-\lambda)v^{\{1\}} + (1-\lambda)v^{\{2\}} + \lambda v^{\{i\}} + \lambda v^{\{j\}} \ . \end{aligned}$$

□

Lemma 6: $\lambda = 1$.

Proof: Consider the set $\{3,4,5\}$. By definition of $v^{A,B}$,

$$v^{\{3,4,5\}} = v^{\{3,4,5\},\{2,5\}} + v^{\{2,5\}} \ .$$

By Lemma 3 the right hand side equals

$$\lambda v^{\{3,4\},\{2\}} + v^{\{2,5\}} = \lambda v^{\{3,4\}} - \lambda v^{\{2\}} + v^{\{2,5\}}$$

where the last equality follows from $v^{A,B} = v^A - v^B$.

Using Lemma 5 for the set $\{3,4\}$ and Lemma 4 for the set $\{2,5\}$, we get

$$\begin{aligned} v^{\{3,4,5\}} &= \lambda[(1-\lambda)v^{\{1\}} + (1-\lambda)v^{\{2\}} + \lambda v^{\{3\}} + \lambda v^{\{4\}}] - \lambda v^{\{2\}} \\ &\quad +(1-\lambda)v^{\{1\}} + v^{\{2\}} + \lambda v^{\{5\}} \\ &= (1-\lambda^2)v^{\{1\}} + (1-\lambda^2)v^{\{2\}} + \lambda^2 v^{\{3\}} + \lambda^2 v^{\{4\}} + \lambda v^{\{5\}} \end{aligned}$$

By symmetry between 4 and 5, we also get

$$v^{\{3,4,5\}} = (1-\lambda^2)v^{\{1\}} + (1-\lambda^2)v^{\{2\}} + \lambda^2 v^{\{3\}} + \lambda v^{\{4\}} + \lambda^2 v^{\{5\}}.$$

Equating the two, it has to be the case that

$$\lambda v^{\{4\}} + v^{\{5\}} = v^{\{4\}} + \lambda v^{\{5\}}$$

or

$$(1-\lambda)v^{\{4\},\{5\}} = 0 \ .$$

By the diversity axiom, $v^{\{4\},\{5\}} \neq 0$, hence the conclusion follows. □

Lemma 7: *For every* $i, j \in \Omega$, $v^{\{i,j\}} = v^{\{i\}} + v^{\{j\}}$.

Proof: Use Lemmata 4,5, and 6. □

Lemma 8: *For every* $A \in \Sigma'$, $v^A = \sum_{i \in A} v^{\{i\}}$.

Proof: By induction on $|A|$. We already know that the lemma holds for $|A| \leq 2$. Assume it is true for $|A| \leq k$ and consider a set B with $|B| = k+1 < n-1$. Choose $i \in B$ and $j \notin B$. Since B and $\{i,j\}$ are non-included, we may write

$$v^B = v^{B,\{i,j\}} + v^{\{i,j\}} \ .$$

Denoting $A = B\backslash\{i\}$, we can write $v^{B,\{i,j\}} = v^{A\cup\{i\},\{j\}\cup\{i\}}$ where $i \notin A \cup \{j\}$. Using lemmata 3 and 6,

$$v^{B,\{i,j\}} = v^{A\cup\{i\},\{j\}\cup\{i\}} = v^{A,\{j\}}$$

plugging this into the first equation we get

$$v^B = v^{A,\{j\}} + v^{\{i,j\}} = v^A - v^{\{j\}} + v^{\{i,j\}}$$

using the induction hypothesis

$$= \sum_{k\in A} v^{\{k\}} - v^{\{j\}} + v^{\{i\}} + v^{\{j\}} = \sum_{k\in B} v^{\{k\}} .$$

This completes the proof of Theorem 2.4 for the finite case. $\square$

We now turn to the case of an infinite Ω. Choose an element of Ω, say, 1. Consider all the finite sub-algebras Σ_0 of Σ that include $\{1\}$ and that have at least five atoms. For each such Σ_0 there exists a matrix $v^{\Sigma_0'} : \Sigma_0' \times M \to R$ that ranks Σ_0' and that satisfies additivity with respect to the union of disjoint sets. Further, such a matrix is unique up to multiplication by a positive scalar. Choose one such Σ_0 and a corresponding $v^{\Sigma_0'}$ for it. Let Σ_1 be another sub-algebra of Σ that includes $\{1\}$ and that has at least five atoms. Let $v^{\Sigma_1'}$ be a matrix that ranks Σ_1'. Let Σ_2 be the minimal algebra containing both Σ_0 and Σ_1. Applying the result to the (finite) sub-algebra Σ_2, there exists $v^{\Sigma_2'}$ that ranks Σ_2'. Since $v^{\Sigma_2'}$ ranks both Σ_0' and Σ_1', $\left(v^{\Sigma_2'}(\{1\}, c)\right)_{c\in M}$ differs from both $\left(v^{\Sigma_0'}(\{1\}, c)\right)_{c\in M}$ and $\left(v^{\Sigma_1'}(\{1\}, c)\right)_{c\in M}$ by a positive multiplicative scalar. This implies that the latter two vectors are also differ by a positive multiplicative scalar, and that there is a unique $v^{\Sigma_1'}$ that ranks Σ_1' and agrees with $v^{\Sigma_0'}$ on the row of $\{1\}$. (Observe that this row is not the 0 vector due to the diversity condition.)

Given an event $A \in \Sigma'$, choose a finite sub-algebra Σ_1 as above that includes A, and define $(v(A, c))_{c\in M}$ by the unique $v^{\Sigma_1'}$ identified above. By similar considerations one concludes that this definition of $(v(A, c))_{c\in M}$ does not depend on the choice of Σ_1 that includes A. Hence v is well-defined. Finally, we wish to show that if $A, B, A \cup B \in \Sigma'$, and $A \cap B = \emptyset$, we have

$$v(A \cup B, c) = v(A, c) + v(B, c) \qquad \text{for all } c \in M.$$

Choose a finite sub-algebra Σ_1 as above that includes both A and B, and observe that v coincides with $v^{\Sigma_1'}$ on Σ_1', where $v^{\Sigma_1'}$ satisfies this equality by Lemma 8. □□

Proof of Remark 2.5:

Consider $M = \Omega = \{1, 2, 3, 4\}$ and define $\{\succsim_I\}_I$ by the matrix v given by the table below (where empty entries denote zeroes). It is straightforward to check that $\{\succsim_I\}_I$ satisfies A5 for all triples of events, yet the matrix v is not additive in events.

$v^A(c)$		Case c			
		1	2	3	4
	1	2			
	2		2		
	3			2	
	4				2
Event A	1,2	2	2		
	1,3	2	1	1	
	1,4	2	1		1
	2,3	1	2	1	
	2,4	1	2		1
	3,4	1	1	1	1

Proof of Remark 2.6:

Consider $M = \Omega = \{1, 2, 3, 4, 5\}$ and define $\{\succsim_I\}_I$ as follows: if $A \subseteq B$, then $A \preceq_I B$ for every $I \in \mathbb{J}$. If A is a proper subset of B, then $A \prec_I B$ for every $I \neq 0$. If A and B are non-included, define $\{\succsim_I\}_I$ by the signed measures $\{\mu_c\}_{c \in M}$ given by the table below (where empty entries denote zeroes):

$\mu_c(\{i\})$ Case c

		1	2	3	4	5
	1	-1	1	1	1	1
State i	2	2	1			
	3	4		1		
	4	8			1	
	5	16				1

One may verify that $\{\succsim_I\}_I$ are qualitative probability relations satisfying A2*-A4*, even though $\mu_1(\{1\}) < 0$.

Proof of Theorem 3.1:
We first show

Lemma 1: *The symmetry axiom implies that there are two numbers, $a > b \geq 0$, such that $\mu_i(\{j\}) = b$ if $i \neq j$ and $\mu_i(\{i\}) = a$.*

Observe that this condition is also equivalent to the symmetry axiom.

Proof of Lemma 1:
Claim 1: *For every $i \leq n$ there exists a number $b_i \in \mathbb{R}$, such that $\mu_i(\{j\}) = b_i$ for all $j \neq i$.*
Proof: Consider the memory $I = 1_{\{i\}} \in \mathbb{J}$ and a permutation π that swaps only two states $j, k \neq i$. Since $I = I \circ \pi$, the symmetry axiom implies that $\{j\} \approx_I \{k\}$, hence $\mu_i(\{j\}) = \mu_i(\{k\})$. □

We denote $a_i = \mu_i(\{i\})$.

Claim 2: *For every $i, j \leq n$, $a_j - b_j = a_i - b_i$.*

Proof: Consider $I = 1_{\{i,j\}} \in \mathbb{J}$ and a permutation π that swaps only i, j. Since $I = I \circ \pi$, the symmetry axiom implies that $\{i\} \approx_I \{j\}$, hence $a_j + b_i = b_j + a_i$, or $a_j - b_j = a_i - b_i$. □

Claim 3: *For every distinct $i, j, k \leq n$, $\{k\} \succ_I \{i, j\}$ for $I = 1_{\{k\}} \in \mathbb{J}$.*

Proof: By the diversity axiom we know that there exists a memory $I \in \mathbb{J}$ such that $\{k\} \succ_I \{i,j\}$. The combination axiom implies that there exists $s \leq n$ such that $\{k\} \succ_I \{i,j\}$ for $I = 1_{\{s\}}$. We claim that this can only hold for $s = k$. Indeed, first observe that if $s \notin \{i,j,k\}$, then for $I = 1_{\{s\}}$ we also get, by symmetry, $\{i\} \succ_I \{j,k\}$. But, since $\succsim_I$ is a qualitative probability, $\{k\} \succ_I \{i,j\}$ implies $\{j,k\} \succ_I \{i\}$, a contradiction. Next assume that $s = i$. By similar reasoning, $\{k\} \succ_I \{i,j\}$ implies $\{i,k\} \succ_I \{i,j\}$, but this implies $\{k\} \succ_I \{j\}$, which contradicts symmetry (as in Claim 1). $s = j$ is similarly excluded, and the conclusion follows. □

Claim 4: *For every* $i, j \leq n$, $a_i = a_j > 0$.

Proof: Let there be given distinct $i, j, k \leq n$. Consider, for every two non-negative integers m, l, the memory $I = I(m,l) = m1_{\{j\}} + 1_{\{k\}} \in \mathbb{J}$, and the permutation π that swaps only i, j. Thus, $I \circ \pi = m1_{\{i\}} + l1_{\{k\}}$. It follows from the symmetry axiom that $\{i,j\} \succ_I \{k\}$ iff $\{i,j\} \succ_{I \circ \pi} \{k\}$. This means that, for every $m, l \geq 0$,

$$ma_i + mb_i + 2lb_k \geq mb_i + la_k \quad \text{iff} \quad ma_j + mb_j + 2lb_k \geq mb_j + la_k$$

or

$$ma_i \geq l(a_k - 2b_k) \quad \text{iff} \quad ma_j \geq l(a_k - 2b_k) \ . \ (*)$$

Further, we argue that $a_i, a_j, (a_k - 2b_k) > 0$. First observe that, by Claim 3 (corresponding to $m = 0, l = 1$), it has to be the case that $(a_k - 2b_k) > 0$. Also, Claim 3 implies that $\{i\} \succ_I \{j,k\}$ for $I = 1_{\{i\}}$ (corresponding to $m = 1$, $l = 0$), and $\{i,j\} \succ_I \{k\}$ follows by monotonicity of $\succsim_I$ with respect to set inclusion. Hence $a_i > 0$. Similarly, $a_j > 0$ has to hold as well. The desired result now follows from $(*)$. □

Combining Claims 1, 2, and 4, we conclude that there are two numbers, $a, b \in \mathbb{R}$, such that $\mu_i(\{j\}) = b$ if $i \neq j$ and $\mu_i(\{i\}) = a$. Furthermore, we know that $a > 0$.

Claim 5: $a > b$.

Proof: As in the proof of Claim 4, $(a - 2b) > 0$ and this suffices since $a > 0$. $\square$

Claim 6: $b \geq 0$.

Proof: Let $i, j, k \leq n$ be distinct. Consider $I = 1_{\{i,j\}}$. We know that $\{i\} \approx_I \{j\}$, hence $\{i,k\} \succsim_I \{j\}$. Hence $a + 3b \geq a + b$. $\square$

This completes the proof of Lemma 1. It remains to show that the specificity axiom implies that $b = 0$ as well. To see this, consider again the proof of Claim 6 above, and observe that if $b > 0$, it has to be the case that $\{i,k\} \succ_I \{j\}$, where $I = 1_{\{i,j\}}$. But this is a contradiction to the specificity axiom, because $\{i\} \approx_I \{j\}$ and $I(k) = 0$. $\square\square$

Proof of Remark 3.2:

Assume that for some $i, j \in \Omega$, $\mu_i(\{j\}) < 0$. This means that for $I = 1_{\{i\}}$ we have $\emptyset \succ_I \{j\}$, contradicting A1'. Hence μ_i are non-negative. Next we consider $i \neq j$ and show that $\mu_i(\{j\}) = 0$. Assume, to the contrary, that $\mu_i(\{j\}) > 0$ for $i \neq j$. In this case, for $I = 1_{\{i\}}$ we have $\emptyset \succsim_I \emptyset$ and $\{j\} \succ_I \emptyset$ while $I(j) = 0$, contradicting A7. Also, since, by A1', $\Omega \succ_I \emptyset$ for $I = 1_{\{i\}}$, it follows that $\mu_i(\{i\}) > 0$ for all $i \in \Omega$. Finally, consider $I = 1_{\{i,j\}}$ for $i \neq j$. By A6, $\{i\} \approx_I \{j\}$. Hence $\mu_i(\{i\}) = \mu_j(\{j\})$. $\square\square$

References

[1] Aumann, R. J. (1995), "Interactive Epistemology", Discussion Paper no. 67, Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.

de Finetti, B. (1931), "Sul Significato Soggettivo della Probabilità", *Fundamenta Mathematicae*, **17**: 298-329.

de Finetti, B. (1937), "La Prevision: Ses Lois Logiques, Ses Sources Subjectives", *Annales de l'Institute Henri Poincare*, **7**: 1-68.

Fagin, R. and J. Y. Halpern (1994), "Reasoning about Knowledge and Probability", *Journal of the Association for Computing Machinery* **41**: 340-367.

Fagin, R., J. Y. Halpern, and N. Megiddo (1990), "A Logic for Reasoning about Probabilities", *Inf. Comput.* **87**: 78-128.

Fine, T. (1973), *Theories of Probability.* New York: Academic Press.

Fishburn, P. C. (1986), "The Axioms of Subjective Probability", *Statistical Science,* **1**: 335-358.

Gilboa, I. and D. Schmeidler (1999), "Inductive Inference: An Axiomatic Approach", Foerder Institute for Economic Research Working Paper No. 29-99. Revised, 2001.

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

Heifetz, A., and P. Mongin (1999), "Probability Logic for Type Spaces", *Games and Economic Behavior*, forthcoming.

Kraft, C. H., J. W. Pratt, and A. Seidenberg (1959), "Intuitive Probability on Finite Sets", *Annals of Mathematical Statistics* **30**: 408-419.

Krantz, D. H., D. R. Luce, P. Suppes, and A. Tversky (1971), *Foundations of Measurement I.* New York: Academic Press.

Ramsey, F. P. (1931), "Truth and Probability", *The Foundation of Mathematics and Other Logical Essays*. New York: Harcourt, Brace and Co.

Savage, L. J. (1954), *The Foundations of Statistics*. New York: John Wiley and Sons.