

Elementary Non-Archimedean Representations of Probability for Decision Theory and Games

PETER J. HAMMOND

An earlier October 1988 version entitled ‘Extended Probabilities for Decision Theory and Games’ was presented to the Econometric Society European Meeting in Bologna 1988. An extensive revision was part of a presentation to the Second Workshop on ‘Knowledge, Belief and Strategic Interaction: The Problem of Learning’ at Castiglioncello (Livorno), Italy in June 1992. This August 1992 version is intended as a contribution to a two volume collection honouring Patrick Suppes, to be edited by Paul Humphreys and published by Kluwer Academic Publishers.

ABSTRACT. In an extensive form game, whether a player has a better strategy than in a presumed equilibrium depends on the other players’ equilibrium reactions to a counterfactual deviation. To allow conditioning on counterfactual events with prior probability zero, extended probabilities are proposed and given the four equivalent characterizations: (i) complete conditional probability systems; (ii) lexicographic hierarchies of probabilities; (iii) extended logarithmic likelihood ratios; and (iv) certain ‘canonical rational probability functions’ representing ‘trembles’ directly as infinitesimal probabilities. However, having joint probability distributions be uniquely determined by independent marginal probability distributions requires general probabilities taking values in a space no smaller than the non-Archimedean ordered field whose members are rational functions of a particular infinitesimal.

Elinor now found the difference between the expectation of an unpleasant event, however certain the mind may be told to consider it, and certainty itself.

— JANE AUSTEN, *Sense and Sensibility*, CH. 48.

... a more attractive and manageable theory may result from a non-Archimedean representation. ... One must keep in mind the fact that the refutability of axioms depends both on their mathematical form and their empirical interpretation.

— KRANTZ, LUCE, SUPPES AND TVERSKY (1971, P. 29).

Non-Archimedean Probabilities

1. Introduction and Outline

Patrick Suppes is a philosopher of science in the broadest sense, which includes the social and behavioural sciences. For many years he directed the interdisciplinary Institute of Mathematical Studies in the Social Sciences at Stanford. He has displayed an abiding interest in the probabilistic models that are used in many branches of science, including physics and quantum mechanics, psychology and linguistics, as exemplified by Suppes (1984) and the work cited therein. Some of his early work was for the Stanford Value Theory Project, producing publications in Bayesian decision theory and subjective probability such as Davidson, McKinsey and Suppes (1955), Davidson and Suppes (1956), Suppes (1956), Davidson, Suppes and Siegel (1955) — see also Suppes (1961), Luce and Suppes (1965), and other papers collected in Part II of Suppes (1969). This is the theme that will be revisited here and in later closely related work.

Krantz, Luce, Suppes and Tversky also collaborated over many years in a large project on the foundations of measurement, resulting in the three volume work that I shall refer to as KLST. A crucial assumption in many of the systems of measurement they consider is the Archimedean axiom. For the case of an ordered algebraic field such as the real line, this axiom requires that, whenever q is a positive quantity, there exists some integer n for which $nq > 1$. In the case where q is infinitesimal, this axiom is violated. KLST discuss variations of this axiom in volume I (Sections 1.4.4, 1.5.2) and in volume III (Section 19.2.3). Furthermore, in volume III (Section 21.7) they argue that one is free to impose the Archimedean axiom or not on most structures of measurement (see also Robinson, 1951 and Narens, 1974a). An analogy is Cohen's demonstration that one is free to include or exclude the axiom of choice from Zermelo-Fraenkel set theory, without fear of contradiction in either case, as explained by Suppes (1972, pp. 250 and 252). Yet KLST chose not to pursue the topic of non-Archimedean measurement in any depth, instead referring the reader to the work of others. Narens (1985) in particular extends some of KLST's ideas to measures taking non-standard values, including infinitesimals and infinite numbers. Very recently, however, Suppes himself has been collaborating with Rolando Chuaqui on other ideas making use of non-standard analysis.

In the case of Bayesian decision theory and the associated theory of games, I am going to argue that non-Archimedean representations of probability are absolutely essential if one is going to avoid some crucial difficulties arising from the need to condition on events that have probability zero. To see why this is important, Section 2 provides two examples showing how easily paradoxes arise in extensive form games.

To avoid such paradoxes by allowing conditioning on events having probability zero, several different kinds of ‘extended probabilities’ have been proposed in the past. So, after some necessary preliminaries, Section 3 begins with a brief history of conditional probability systems. Thereafter it gives formal definitions of: (i) complete conditional probability systems, as considered by Rényi (1955), Myerson (1986), and others; (ii) lexicographic hierarchies of probabilities with disjoint supports, as considered recently by Blume, Brandenburger and Dekel (1991a, b) [henceforward frequently referred to as BBD]; and (iii) extended logarithmic likelihood ratios, as considered by McClellan (1989a, b). BBD also made use of a fourth and richer set of lexicographic hierarchies of probabilities with supports that may not be disjoint, which is defined as well.

Thereafter Section 4 shows that spaces (i)–(iii) above are just three different but equivalent versions of one main or ‘canonical’ space of extended probabilities. By constructing suitable metrics, the three can even be made into homeomorphic compact metric spaces.

In decision trees with uncertainty, the random moves that occur at different chance nodes ought to be independent because any cause of dependence ought to be modelled within the tree. Moreover, knowing the independent probability distributions at different chance nodes should be enough to determine the entire joint distribution of random moves in the entire decision tree. While this is all elementary in decision trees with ordinary probabilities at each chance node, some new conceptual issues need to be resolved when extended probabilities are being considered. Accordingly, Section 5 considers joint distributions of pairs of independent random variables. Two different definitions of independence are considered — namely, almost sure independence, and conditional independence. The second is more restrictive than the first. However, neither passes the test of allowing the joint distribution to be inferred from the appropriate independent marginal distributions.

In conventional probability theory, the joint distribution of independent random variables can be constructed by simply multiplying the marginal distributions. In order that

extended probabilities should enjoy the same property, it is argued that they should also be given values in an algebraic field. Only an algebraic field is closed under multiplication and division (except by zero), as well as under addition and subtraction. Section 6 therefore begins by presenting the simplest possible non-Archimedean ordered field that includes the real line \mathfrak{R} . This is $\mathfrak{R}(\epsilon)$, defined as the smallest field containing both \mathfrak{R} and the single positive infinitesimal denoted by ϵ . ‘Elementary’ non-Archimedean probability distributions can then be regarded as having values in $\mathfrak{R}(\epsilon)$ which are determined by ‘rational probability functions’ evaluated at ϵ instead of at a real value. Such probabilities represent directly the kind of trembles considered previously by Selten (1975) and Myerson (1978). Here and in later work such as Hammond (1992), it will be argued that no smaller set of probabilities is adequate for Bayesian decision theory when counterfactuals need to be considered. It is also shown that the extended probability spaces of Section 3 are equivalent to proper subsets of this space, so that none of them is rich enough for analysing decision trees or extensive games. Moreover, whereas ‘canonical’ extended probabilities give rise to real valued conditional probabilities, for rational probability functions even the corresponding conditional probabilities are generally non-Archimedean.

Section 7 has a brief concluding summary and points ahead to later work making use of the non-Archimedean probabilities that are introduced in this paper.

2. Motivation: The Zero Probability Problem

2.1. A Team Example

In any equilibrium of a game in extensive form, the probability attached to each deviation from an equilibrium strategy is zero. Yet in order to tell whether a player has a better strategy than in a presumed equilibrium, the reactions of the other players to this deviation, as prescribed by their equilibrium strategies, must be known. Since the deviation has probability zero there is no guarantee that, even in equilibrium, these reactions will be either sensible or credible behaviour at those information sets which are reached with probability zero in equilibrium. This point has been made especially clearly and forcefully by Selten (1965, 1973, 1975).

A simple example of the problem is the extensive form game illustrated in Figure 1. Player I has the move at node n_0 and player II at n_1 . This is a ‘team’ game, in the sense

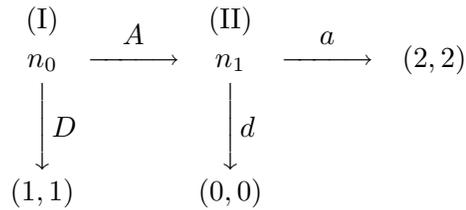


Figure 1

of Marschak and Radner (1972), with the two players having identical payoffs at all three terminal nodes. The set of Nash equilibria is easily computed to consist of:

- (i) the pure strategy pair (A, a) ;
- (ii) any strategy pair of the form $(D, \alpha a + (1 - \alpha)d)$ for $0 \leq \alpha \leq 1/2$, in which player I chooses a pure strategy and II may have a mixed strategy.

In particular, all the equilibria of type (ii) involve II playing the weakly dominated strategy d with positive probability. Such equilibria are also ‘subgame imperfect,’ in the standard terminology due to Selten (1973). Indeed, the example is very similar to the one originally used by Selten (1965) to motivate subgame perfect equilibria.

To my mind, however, this subgame imperfection manifests a more fundamental problem — namely, that prescribed behaviour ought to be ‘dynamically consistent’ in the following sense (cf. Hammond, 1988a, b). Regardless of whether n_1 is viewed as a decision node of the whole tree illustrated in Figure 1, or merely of the subtree whose initial node is n_1 , the ‘behaviour set’ that is prescribed for player II — which is of course a non-empty subset of $\{a, d\}$ — should be the same. Yet it is absurd to prescribe $\{d\}$ in the subtree, or to countenance II using any mixed strategy which gives positive probability to strategy d . So in the game illustrated dynamic consistency rules out using any mixed strategy giving positive probability to strategy d .

The example in Figure 1 also illustrates the difficulty in applying Bayesian updating to probabilistic beliefs in extensive form games. Take the pure strategy Nash equilibrium of type (ii), namely (D, d) . It is a Nash equilibrium because the probability of reaching n_1 , where player II has a move, is zero, and so it does not matter that d is not II’s optimal strategy in the subgame. In fact, given the degenerate distribution that attaches probability one to the players’ choosing the equilibrium strategies (D, d) , the probabilities of d and a

conditional on player I's counterfactual choice of A instead cannot be determined just from Bayesian updating. Nor can II's expected payoff conditional on player I choosing A instead of D . Yet there really are well-defined payoffs for player II in the subgame. Thus, given these subgame payoffs, the conditional probability of d given A should really be zero. So I's expected payoff, conditional on deviating to A , should really be 2. This implies that A is player I's unique optimal strategy, and so excludes the subgame imperfect Nash equilibrium (D, d) .

To conclude, Bayesian updating breaks down in any subgame that is reached counterfactually, with probability zero, in Nash equilibrium. Instead, some other way has to be found of circumventing zero probabilities so that player II's conditional expected payoffs are well defined, and so that player I is forced to have realistic conditional beliefs about what II's strategic reaction would be if player I were to choose A after all. Sections 3–5 below discuss the several kinds of extended probability with which game theorists have sought to remedy this problem. Section 6 will propose using a particular set of infinitesimal probabilities.

2.2. A Single Person Example

The example of Figure 1 had two players — albeit two players with identical payoffs. A similar problem arises even in single person decision theory, however, as is illustrated by the simple one person game shown in Figure 2.

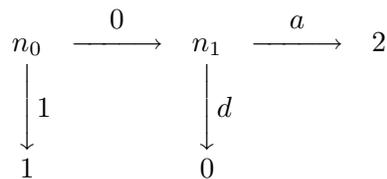


Figure 2

This is just a one-person version of the team game shown in Figure 1. The initial node n_0 of the decision tree is a chance node at which there is a probability one of making the transition down to the terminal node below where the agent's payoff is 1, and there is a probability zero of making the transition across to the decision node n_1 on the right. At n_1 the agent has the choice of moving to the terminal node on the right where the agent's payoff is 2, or of moving down to the terminal node below where the payoff is 0. In the

normal form of this game the agent has two strategies a and d , each of which yields the same expected payoff 1. Thus any mixture of the two strategies a and d is an equilibrium of this game in normal form. Yet dynamic consistency implies that the same behaviour — any mixture of the two strategies a and d — should be an equilibrium in the subtree which starts with the decision node n_1 . In this subtree, however, the agent effectively confronts an even simpler decision problem: the choice between a with payoff 2 and d with payoff 0. The only appropriate equilibrium behaviour, of course, involves choosing a for sure. So there is no way of obtaining dynamically consistent behaviour if one considers just the normal form of the one person game illustrated in Figure 2.

One might argue, as I did in Hammond (1988a, b), that the example of Figure 2 is easily dealt with by excluding from each single person decision tree the result of any chance move which occurs with zero probability. Then the tree of Figure 2, for example, would be pruned so that it contains only the initial node n_0 , giving rise directly to the payoff 1. There is no need to contemplate what decision should be made at n_1 until after a zero probability event has actually occurred. This simple way out of the difficulty is no longer available, however, when credible equilibria of multi-person games are being considered. Indeed, Figure 1 illustrates how the same difficulty arises even in team games where there is perfect information. In that example, it is not appropriate for player I simply to choose D and then ignore what happens in the subgame that is reached with zero probability.

3. Four Spaces of Extended Probabilities

3.1. Preliminaries

Let Ω be the non-empty *sample space* Ω , which may be finite or infinite. Let $\Delta(\Omega)$ denote the set of *simple probability measures* $p(\cdot)$ on the set Ω — i.e., the set of additive functions $p(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1] \subset \Re$ which are defined on the power set $\mathcal{P}(\Omega)$ of all subsets of Ω and which, for some non-empty finite set $F \subset \Omega$, satisfy the two restrictions that

$$p(E) = p(E \cap F) \quad (\text{all } E \subset \Omega) \quad \text{and} \quad p(F) = 1. \quad (1)$$

It loses no generality to suppose that F has been chosen to satisfy $p(\omega) > 0$ for all $\omega \in F$, in which case F is called the *support* of $p(\cdot)$.

Suppose that F is a non-empty and finite subset of the sample space Ω . Notice that the set $\Delta(F)$ is equivalent to the well defined subset of $\Delta(\Omega)$ whose members satisfy (1). Moreover, $\Delta(F)$ can be regarded as consisting of the *conditional* probability measures $P(\cdot|F)$ on F which satisfy $P(F|F) = 1$. Also, let $\Delta^0(F)$ denote the subset of *interior* probability measures $p(\cdot) \in \Delta(F)$ on F satisfying $p(\omega) > 0$ for all $\omega \in F$. Thus $\Delta^0(F)$ consists of those measures in $\Delta(F)$ whose support is the whole set F .

3.2. Complete Conditional Probability Systems

An obvious attempt to remedy the zero probability problem is to treat conditional probabilities as the basis of probability theory. As Rényi (1964) points out, this idea is actually quite old, going back to works such as Keynes (1921), Koopman (1940), Barnard (1949), and Good (1950). Some early formal treatments can be found in Popper (1934, 1938) and de Finetti (1936, 1949). The latter explicitly allowed conditioning on events with probability zero, but did so without necessarily invoking σ -additivity.

Kolmogorov's (1933) standard measure-theoretic framework was later generalized by Rényi (1955, 1956, 1964, 1970) and Császár (1955), who introduced 'conditional probability spaces' with a σ -algebra of events and also an algebra of 'conditioning events.' For a rather different axiomatization, see Krantz, Luce, Suppes and Tversky (1971, pp. 220–8). From the point of view of this paper, however, the key contribution came from allowing some of the conditioning events to have probability zero. Of course, it is common for continuous density functions to be conditioned on zero probability events — for a formal discussion of some of the problems then raised, see Blackwell and Ryll-Nardzewski (1963) and Blackwell and Dubins (1975). Here, such conditioning occurs even for discrete distributions, as in Lindley's (1965, especially p. 6) elementary textbook presentation.

In fact, Popper's (1938) earlier article was explicitly designed to lay out a system of axioms far more general than those of Kolmogorov or even Rényi — see Popper (1959, pp. 318–348). 'Extended' conditional probabilities have also been discussed by other philosophers such as Stalnaker (1970), Lewis (1973), Harper (1975), and Levi (1980) who, following Popper, were concerned with counterfactuals and conditional logic — see also the collection of papers in Harper, Stalnaker and Pearce (1981). Some work in a similar vein, but with game theory explicitly in mind, is due to Selten and Leopold (1982).

What Rényi and Czászár called ‘conditional probability spaces’ were introduced into game theory by Myerson (1986) under the name of ‘complete conditional probability systems’ (or CCPS’s for short). They have since been used by Fudenberg and Tirole (1991), amongst others. Such systems specify probabilities conditional even on zero probability events. Following an idea due to Kreps and Wilson (1982, p. 874), Myerson defined this set of CCPS’s as the closure of the set of usual conditional probabilities derived from the interior probability distributions. The definition given below is more direct than Myerson’s, and is equivalent to Lindley’s (1965, p. 6).

First, if F is any non-empty finite subset of the sample space Ω , the domain of event pairs for which conditional probabilities ought to be defined is

$$\mathcal{E}(F) := \{ (E, E') \mid \emptyset \neq E \subset E' \subset F \}. \quad (2)$$

A *complete conditional probability system* (CCPS) on F is a mapping $P(\cdot|\cdot) : \mathcal{E}(F) \rightarrow [0, 1]$ that defines conditional probabilities $P(E_1|E_2)$ for all $(E_1, E_2) \in \mathcal{E}(F)$ which satisfy the conditions

$$P(\cdot|E) \in \Delta(E) \quad (\text{all non-empty } E \subset F); \quad (3)$$

$$P(E_1|E_3) = P(E_1|E_2)P(E_2|E_3) \quad (\text{all non-empty } E_1 \subset E_2 \subset E_3 \subset F). \quad (4)$$

When $P(E_2|E_3) \neq 0$ the restrictions (4) are equivalent to Bayes’ rule that

$$P(E_1|E_2) = P(E_1|E_3)/P(E_2|E_3). \quad (5)$$

Unlike (5), however, (4) is also valid and necessary even when $P(E_2|E_3) = 0$.

Let $\Delta_C(F)$ denote the set of all such CCPS’s on the finite set F . Evidently, for any given CCPS $P(\cdot|\cdot) \in \Delta_C(F)$, there exists a unique corresponding probability distribution $p(\cdot) \in \Delta(F)$ given by $p(E) = P(E|F)$ (all non-empty $E \subset F$). Moreover, when $P(E_1|E_2) > 0$ everywhere in $\mathcal{E}(F)$ it is obvious that $p(\cdot) \in \Delta^0(F)$. Conversely, for any interior $p(\cdot) \in \Delta^0(F)$ there will be a unique corresponding CCPS $P(\cdot|\cdot) \in \Delta_C(F)$ defined for all $(E_1, E_2) \in \mathcal{E}(F)$ by the obvious equation

$$P(E_1|E_2) = p(E_1)/p(E_2). \quad (6)$$

But this definition evidently fails in the more troublesome case when $p(\cdot) \notin \Delta^0(F)$ and moreover there is a subset E of F with at least two distinct members for which $p(E) = 0$.

There is one remaining marginal case when $p(\cdot) \notin \Delta^0(F)$, but when there also exists a unique state $\omega \in F$ such that $p(\omega) = 0$. Then (6) determines $P(E_1|E_2)$ throughout $\mathcal{E}(F)$ except when $E_1 = E_2 = \{\omega\}$, in which case one must of course take $P(\{\omega\}|\{\omega\}) = 1$.

3.3. Lexicographic Conditional Probability Systems

Lexicographic utilities were originally introduced by Hausner (1954), Thrall (1954). Thereafter they were considered further by Chipman (1960, 1971a, 1971b), Richter (1971) and Skala (1975). However, where these works consider lexicographic expected utility at all, it is as the expected value of lexicographic utility with respect to ordinary probabilities, rather than as the expected value of ordinary utility with respect to lexicographic probabilities. Chernoff (1954, pp. 440–1), on the other hand, did provide an example of a decision criterion which is equivalent to a form of lexicographic expected utility maximization and is related to the use of infinitesimal probabilities. Much more systematic were the articles of Rényi (1955, 1956) and Császár (1955). They proved the existence of a ‘dimensionally ordered set of measures’ — i.e., a (possibly infinite) lexicographic hierarchy of probability measures — which can be used to represent conditional probabilities when some of the conditioning events are allowed to have probability zero. More recently, in a non-probabilistic framework, Spohn (1988, 1990) has used a hierarchy of increasingly exceptional propositions to represent some important aspects of non-monotonic reasoning.

In their work on sequential equilibrium in extensive form games, Kreps and Wilson (1982) needed to describe beliefs conditional on reaching an information set which occurs with probability zero. To do so they considered the ‘lexicographic consistency’ of a hierarchy of probabilistic hypotheses. This idea has been taken much further by BBD, who give a systematic treatment of general lexicographic hierarchies of subjective probability distributions. They show how such hierarchies can be used to justify Myerson’s (1978) concept of proper equilibrium for games in normal form, as well as Selten’s (1975) ‘trembling hand perfect’ equilibria. The formulation below is somewhat different because this paper discusses variable objective probabilities, rather than a single fixed subjective probability distribution over uncertain states of the world.

Formally, let F be any finite non-empty subset of the sample space Ω . Then a *lexicographic probability system* (or LPS) on F is any ordered finite collection or *hierarchy*

$\mathbf{p} = \langle p_k \rangle_{k=0}^K$ of probability distributions $p_k(\cdot) \in \Delta(F)$ satisfying the restriction that, for each $\omega \in F$, there exists at least one integer k in the range $k = 0$ to K for which $p_k(\omega) > 0$. In other words, every state $\omega \in F$ is given positive probability by at least one distribution in the hierarchy. Equivalently, it must be true that $\sum_{k=0}^K p_k(\omega) > 0$. Let $\Delta_L^*(F)$ denote the set of all such LPS's.

The intended interpretation of any such hierarchy is that p_k takes infinite precedence over $p_{k'}$ whenever $k < k'$. It should be noted that BBD (p. 66) did not impose the restriction that $\sum_{k=0}^K p_k(\omega) > 0$ for every $\omega \in F$. But they did observe that it would be satisfied for events which are not 'Savage null' in their theory of subjective probability. In Hammond (1988b, 1992) I argue that no events should be Savage null anyway.

BBD also suggested restricting further the set of LPS's by requiring the distributions $p_k(\cdot) \in \Delta(F)$ ($k = 0$ to K) to have pairwise *disjoint supports* $F_k = \{\omega \in F \mid p_k(\omega) > 0\}$, which therefore partition the set F . This additional restriction turns out to be important, and those LPS's which satisfy it are called *lexicographic conditional probability systems* (or LCPS's) (BBD, p. 71). The space of all such LCPS's will be denoted by $\Delta_L(F)$.

At first, the disjoint supports requirement seems only natural. Any $\omega \in F_k$ must be given positive probability by the k -th member of the LPS hierarchy, so there seems no good reason to include it in any other $F_{k'}$ that has $k' > k$. This intuition turns out to be inadequate, however, as will be seen later in Section 5 and in later work. Moreover, BBD (p. 89) show how imposing disjoint supports may exclude all possible 'lexicographic' Nash equilibria from a game in normal form.

It will turn out that there is a one-to-one correspondence between $\Delta_C(F)$ and $\Delta_L(F)$. This will be demonstrated in Section 4 below, making use of a third kind of extended probability that is introduced next.

3.4. Logarithmic Likelihood Ratio Functions

The last set of extended probabilities to be considered is McLennan's (1989a) space of 'consistent conditional systems.' These specify, for all possible pairs $\omega, \omega' \in F$, the logarithm of the likelihood ratio. Thus, given the orthodox probability distribution $p(\cdot) \in \Delta(F)$, one should have

$$\mu(\omega, \omega') = \ln[p(\omega)/p(\omega')] \tag{7}$$

unless either $p(\omega)$ or $p(\omega')$ is zero. But $\mu(\omega, \omega')$ is also allowed to have one of the two ‘extended’ real values $-\infty$ and $+\infty$, so that (7) is a valid definition unless $p(\omega) = p(\omega') = 0$. Thus $\mu(\omega, \omega')$ will be either $-\infty$ or $+\infty$ if just one of the two probabilities $p(\omega)$ and $p(\omega')$ is zero. But $\mu(\omega, \omega')$ cannot be determined from $p(\omega)$ and $p(\omega')$ alone if both are zero, in which case it becomes necessary to consider the conditional probabilities of ω and ω' given the zero probability event $\{\omega, \omega'\}$. These conditional probabilities, of course, are not defined for orthodox probability distributions, but must be for the extended probabilities which are the subject of this Section.

For any $\omega, \omega', \omega'' \in F$, and whenever all the logarithmic likelihood ratios are well-defined by (7), the standard properties of logarithms imply immediately that

$$\mu(\omega, \omega) = 0; \mu(\omega, \omega') + \mu(\omega', \omega) = 0; \mu(\omega, \omega') + \mu(\omega', \omega'') + \mu(\omega'', \omega) = 0. \quad (8)$$

But even if some of these logarithmic likelihoods are not well-defined by (7), the restrictions (8) are still imposed, naturally enough. Note that the second and third equations of (8) should embody the convention that

$$-\infty + \infty + x = 0 \text{ for all } x \in [-\infty, +\infty]. \quad (9)$$

Formally, let \mathfrak{R}^* denote the extended real line $[-\infty, +\infty]$. Then, given any finite subset F of the sample space Ω , a *logarithmic likelihood ratio function* (or LLRF) is a mapping $\mu(\cdot, \cdot) : F \times F \rightarrow \mathfrak{R}^*$ which is defined for all ordered pairs $(\omega, \omega') \in F \times F$, and which also satisfies the linear restrictions (8), interpreted according to the convention (9) where necessary. The space $\Delta_M(F)$ is defined as the set of all such LLRF’s. This space is linear, as the closure in $(\mathfrak{R}^*)^{F \times F}$ of a linear subspace of $\mathfrak{R}^{F \times F}$. Since $\Delta^0(F)$ is evidently $\#F - 1$ -dimensional, and is isomorphic to a dense subset of $\Delta_M(F)$, it follows that $\Delta_M(F)$ must be a subspace of dimension $\#F - 1$.

4. Equivalence and Homeomorphism

4.1. An Equivalence Theorem

Three natural mappings

$$\psi_{LC} : \Delta_L(F) \rightarrow \Delta_C(F), \quad \psi_{CM} : \Delta_C(F) \rightarrow \Delta_M(F), \quad \psi_{ML} : \Delta_M(F) \rightarrow \Delta_L(F) \quad (10)$$

will now be defined. It will then be shown that these mappings have the property that the threefold composition $\psi_{ML} \circ \psi_{CM} \circ \psi_{LC}$ is the identity mapping from $\Delta_L(F)$ into itself. So, given any member of any one of the three spaces $\Delta_L(F)$, $\Delta_C(F)$, and $\Delta_M(F)$, there exist unique corresponding members of each of the other two. Thus, of the four spaces of extended probabilities introduced in Section 3, the three smallest are actually equivalent.

First, the mapping ψ_{LC} will be defined. For any hierarchy $\mathbf{p} = \langle p_k \rangle_{k=0}^K \in \Delta_L(F)$ and any non-empty $E \subset F$, let $k_{\mathbf{p}}(E) := \min \{ k \mid p_k(E) > 0 \}$. Now suppose that $\emptyset \neq E \subset E' \subset F$. Evidently $k_{\mathbf{p}}(E) \geq k_{\mathbf{p}}(E')$. So define $P(E|E') = \psi_{LC}(\mathbf{p})(E|E')$ to satisfy

$$P(E|E') = \begin{cases} 0 & \text{if } k_{\mathbf{p}}(E) > k_{\mathbf{p}}(E'); \\ p_k(E)/p_k(E') & \text{if } k_{\mathbf{p}}(E) = k_{\mathbf{p}}(E') = k. \end{cases} \quad (11)$$

Then it is routine to check that $P(\cdot|\cdot)$ satisfies both (3) and (4), so that $\psi_{LC}(\mathbf{p}) \in \Delta_C(F)$.

Second, the mapping ψ_{CM} will be defined so that, for any CCPS $P(\cdot|\cdot) \in \Delta_C(F)$ and any pair $\omega, \omega' \in F$, one has

$$\mu(\omega, \omega') = \psi_{CM}(P)(\omega, \omega') = \ln \left[\frac{P(\{\omega\}|\{\omega, \omega'\})}{P(\{\omega'\}|\{\omega, \omega'\})} \right]. \quad (12)$$

Now, for all triples $E = \{\omega, \omega', \omega''\} \subset F$, condition (4) and definition (12) imply that

$$\ln \left[\frac{P(\{\omega\}|E)}{P(\{\omega'\}|E)} \right] = \ln \left[\frac{P(\{\omega\}|\{\omega, \omega'\})}{P(\{\omega'\}|\{\omega, \omega'\})} \times \frac{P(\{\omega, \omega'\}|E)}{P(\{\omega, \omega'\}|E)} \right] = \mu(\omega, \omega') \quad (13)$$

unless $P(\{\omega, \omega'\}|E) = 0$ and so $P(\{\omega''\}|E) = 1$. Then it is trivial to check that (8) holds except when there exists at least one state $\omega^* \in E$ for which $P(\{\omega^*\}|E) = 0$. Even in this case, however, there must still be some extended real number $x \in [-\infty, +\infty]$ such that

$$\mu(\omega, \omega') + \mu(\omega', \omega'') + \mu(\omega'', \omega) = -\infty + \infty + x \quad (14)$$

which must be zero because of the convention (9). Therefore (8) is true in every case, and so $\mu = \psi_{CM}(P)$ must indeed be an LLRF belonging to the space $\Delta_M(F)$.

Third, the mapping $\psi_{ML} : \Delta_M(F) \rightarrow \Delta_L(F)$ will be defined in several stages. To begin with, for any given LLRF $\mu \in \Delta_M(F)$, define the binary relation \succeq on F by

$$\omega \succeq \omega' \iff \mu(\omega, \omega') > -\infty. \quad (15)$$

Note that \succeq is complete and transitive because of the restrictions (8). So one can construct the hierarchy S_k, F_k ($k = 0, 1, 2, \dots$) recursively, starting with $S_0 := F$, and then setting

$$F_k := \{\omega \in S_k \mid \omega' \in S_k \implies \omega \succeq \omega'\}; \quad S_{k+1} := S_k \setminus F_k. \quad (16)$$

For $k = 0, 1, 2, \dots$, the constructed sets must satisfy $S_{k+1} \subset S_k$ and also, because \succeq is complete and transitive, $S_k \setminus S_{k+1} = F_k \neq \emptyset$ whenever $S_k \neq \emptyset$. Because F is finite and $F = S_0 \supset S_1 \supset S_2 \supset \dots \supset S_k \supset S_{k+1} \dots$, it follows that $F_K = S_K$ and so $S_{K+1} = \emptyset$ for some finite K . Note too that the sets $F_k = S_k \setminus S_{k+1}$ ($k = 0$ to K) are pairwise disjoint, and so form a finite partition of F . Now define the unique hierarchy $\langle p_k(\cdot) \rangle_{k=0}^K = \psi_{ML}(\mu)$ corresponding to $\mu \in \Delta_M(F)$ so that, for all $k = 0$ to K and $\omega \in F$, one has

$$p_k(\omega) = \begin{cases} 0 & \text{if } \omega \notin F_k; \\ 1 / \sum_{\omega' \in F_k} \exp[\mu(\omega', \omega)] & \text{if } \omega \in F_k. \end{cases} \quad (17)$$

Because $\mu(\omega', \omega) < +\infty$ for all $\omega', \omega \in F_k$, it follows that $p_k(\omega) > 0$ for all $\omega \in F_k$. Moreover, given any fixed $\bar{\omega} \in F_k$, together (17) and (8) imply that, for all $\omega \in F_k$, one must have

$$\begin{aligned} \frac{1}{p_k(\omega)} &= \sum_{\omega' \in F_k} \exp[\mu(\omega', \omega)] = \sum_{\omega' \in F_k} \exp[\mu(\omega', \bar{\omega}) + \mu(\bar{\omega}, \omega)] \\ &= \exp[\mu(\bar{\omega}, \omega)] \sum_{\omega' \in F_k} \exp[\mu(\omega', \bar{\omega})]. \end{aligned} \quad (18)$$

Now (18) and (8) imply that

$$p_k(\omega) = \exp[\mu(\omega, \bar{\omega})] / \sum_{\omega' \in F_k} \exp[\mu(\omega', \bar{\omega})]. \quad (19)$$

Hence $\sum_{\omega \in F_k} p_k(\omega) = 1$, confirming that $p_k(\cdot) \in \Delta^0(F_k)$. Thus $\psi_{ML}(\mu)$ is indeed an LCPS in the space $\Delta_L(F)$, and so $\psi_{ML} : \Delta_M(F) \rightarrow \Delta_L(F)$.

Finally, it must be shown that the threefold composition $\psi_{ML} \circ \psi_{CM} \circ \psi_{LC}$ is the identity mapping from $\Delta_L(F)$ into itself. Starting with any LCPS $\mathbf{p} = \langle p_k \rangle_{k=0}^K \in \Delta_L(F)$, the corresponding CCPS $P(\cdot|\cdot) = \psi_{LC}(\mathbf{p}) \in \Delta_C(F)$ is given by (11). For each $\omega \in F$,

let $k_{\mathbf{p}}(\omega)$ denote the unique integer k for which $p_k(\omega) > 0$. Because of (12) and (13), the unique corresponding LLRF $\mu = \psi_{CM} \circ \psi_{LC}(\mathbf{p}) \in \Delta_M(F)$ must be such that, whenever $\omega, \omega' \in F$, then

$$\mu(\omega, \omega') = \ln \left[\frac{P(\{\omega\}|\{\omega, \omega'\})}{P(\{\omega'\}|\{\omega, \omega'\})} \right] = \begin{cases} +\infty, & \text{if } k_{\mathbf{p}}(\omega) < k_{\mathbf{p}}(\omega'); \\ \ln[p_k(\omega)/p_k(\omega')], & \text{if } k_{\mathbf{p}}(\omega) = k_{\mathbf{p}}(\omega') = k; \\ -\infty, & \text{if } k_{\mathbf{p}}(\omega) > k_{\mathbf{p}}(\omega'). \end{cases} \quad (20)$$

Next, we find the unique LCPS \mathbf{q} which results from applying the mapping $\psi_{ML} \circ \psi_{CM} \circ \psi_{LC}$ to the LCPS \mathbf{p} . Because of (20), the corresponding ordering of states $\omega \in F$ given by (15) must satisfy

$$\omega \succeq \omega' \iff \mu(\omega, \omega') > -\infty \iff k_{\mathbf{p}}(\omega) \leq k_{\mathbf{p}}(\omega'). \quad (21)$$

For $k = 0$ to K , (20) and (21) evidently imply that the construction (16) leads to

$$\begin{aligned} F_k &= \{ \omega \in S_k \mid \omega' \in S_k \implies k_{\mathbf{p}}(\omega) \leq k_{\mathbf{p}}(\omega') \} = \arg \min_{\omega} \{ k_{\mathbf{p}}(\omega) \mid \omega \in S_k \}; \\ S_{k+1} &= S_k \setminus F_k = \{ \omega \in S_k \mid \exists \omega' \in S_k : k_{\mathbf{p}}(\omega) > k_{\mathbf{p}}(\omega') \}. \end{aligned} \quad (22)$$

But then it follows by induction on k that

$$F_k = \{ \omega \in F \mid k_{\mathbf{p}}(\omega) = k \} \quad \text{and} \quad S_{k+1} = F \setminus \bigcup_{j=1}^k F_j = \{ \omega \in F \mid k_{\mathbf{p}}(\omega) > k \} \quad (23)$$

for $k = 0$ to K . However, for all $\omega \in F_k$, the construction (17) and the equation (20) above together imply that $\mathbf{q} = \psi_{ML}(\mu)$ must satisfy

$$\frac{1}{q_k(\omega)} = \sum_{\omega' \in F_k} \exp[\mu(\omega', \omega)] = \sum_{\omega' \in F_k} \frac{p_k(\omega')}{p_k(\omega)} = \frac{1}{p_k(\omega)}. \quad (24)$$

Thus, $\psi_{ML} \circ \psi_{CM} \circ \psi_{LC}$ is indeed the identity mapping from $\Delta_L(F)$ into itself. So the three spaces must be equivalent, as claimed.

4.2. Homeomorphic Metric Spaces

A collection of metrics for each of the three spaces of extended probabilities $\Delta_L(F)$, $\Delta_C(F)$, and $\Delta_M(F)$ will now be constructed in a way that creates a homeomorphism between each pair of spaces. This collection will be based upon the particular metric d_C for the space $\Delta_C(F)$ which is defined, for all pairs $P, Q \in \Delta_C(F)$, by

$$d_C(P, Q) := \max_{\omega, \omega' \in F} \{ |P(\{\omega\}|\{\omega, \omega'\}) - Q(\{\omega\}|\{\omega, \omega'\})| \}. \quad (25)$$

After this definition for $\Delta_C(F)$, the metrics on the other two spaces will be constructed to ensure that homeomorphism is automatically satisfied. This involves defining the corresponding metric d_M for $\Delta_M(F)$ by

$$d_M(\lambda, \mu) := d_C(\psi_{CM}^{-1}(\lambda), \psi_{CM}^{-1}(\mu)) = \max_{\omega, \omega' \in F} \left\{ \left| \frac{1}{1 + \exp \lambda(\omega', \omega)} - \frac{1}{1 + \exp \mu(\omega', \omega)} \right| \right\} \quad (26)$$

for all $\lambda, \mu \in \Delta_M(F)$. The corresponding metric d_L for $\Delta_L(F)$ is given by

$$d_L(\mathbf{q}, \mathbf{r}) := d_C(\psi_{LC}(\mathbf{q}), \psi_{LC}(\mathbf{r})) = \max_{\omega, \omega' \in F} \{ |\delta_{\mathbf{q}}(\omega, \omega') - \delta_{\mathbf{r}}(\omega, \omega')| \} \quad (27)$$

for all $\mathbf{q}, \mathbf{r} \in \Delta_M(F)$ where, given any $\mathbf{p} \in \Delta_L(F)$ and any pair $\omega, \omega' \in F$,

$$\delta_{\mathbf{p}}(\omega, \omega') := \psi_{LC}(\mathbf{p})(\{\omega\}|\{\omega, \omega'\}) = \begin{cases} 1 & \text{if } k_{\mathbf{p}}(\omega) < k_{\mathbf{p}}(\omega'); \\ p_k(\omega)/[p_k(\omega) + p_k(\omega')] & \text{if } k_{\mathbf{p}}(\omega) = k_{\mathbf{p}}(\omega') = k; \\ 0 & \text{if } k_{\mathbf{p}}(\omega) > k_{\mathbf{p}}(\omega'). \end{cases}$$

denotes the corresponding conditional probability of ω given $\{\omega, \omega'\}$.

Note first that all the three corresponding functions d_C , d_M , and d_R really are metrics. This is because each satisfies the usual triangle inequality, and because $d_M(\lambda, \mu) = 0$, for instance, implies that $\lambda(\omega, \omega') = \mu(\omega, \omega')$ for all $\omega, \omega' \in F$, so that $\lambda = \mu$ in $\Delta_M(F)$. Then it is obvious from these constructions that the three metric spaces

$$(\Delta_C(F), d_C), (\Delta_M(F), d_M), (\Delta_L(F), d_L) \quad (28)$$

are homeomorphic. Moreover, $(\Delta_C(F), d_C)$ is clearly compact in the finite dimensional Euclidean space $\mathfrak{R}^{\mathcal{E}(F)}$, as a closed subset of the Cartesian product set $[0, 1]^{\mathcal{E}(F)}$. Therefore all three spaces are compact. In fact, it follows from McLennan (1989b) that all three spaces can be made homeomorphic to the closed unit ball in $\mathfrak{R}^{\#F-1}$.

5. Independence and Joint Distributions

5.1. Almost Sure Independence

Consider a sample space in the form of a twofold Cartesian product $\Omega = \Omega^A \times \Omega^B$. Suppose too that the non-empty finite subset $F \subset \Omega$ can be expressed as $F = F^A \times F^B$ where $F^A \subset \Omega^A$ and $F^B \subset \Omega^B$. For example, F^A and F^B could be thought of as the moves which nature might make at two different chance nodes n^A and n^B of a particular decision tree. The latter part of BBD (1991a) distinguishes three different ways in which a joint extended probability distribution on pairs $(\omega^A, \omega^B) \in F^A \times F^B$ might be independent. This section considers two of those ways, or minor variations of them; the third way is the topic of Section 6.5.

Of BBD's three forms of independence, the weakest is when the joint distribution is an 'approximate product measure' (p. 75). This is identical to the following definition. Say that the joint complete conditional probability system $P(\cdot|\cdot) \in \Delta_C(F^A \times F^B)$ is *almost surely independent* if its support is the Cartesian product of two subsets $F_0^A \subset F^A$ and $F_0^B \subset F^B$, and if there exist marginal distributions $P^i(\cdot) \in \Delta(F_0^i)$ ($i = A, B$) for which

$$P(E^A \times E^B | F_0^A \times F_0^B) = P^A(E^A) \times P^B(E^B)$$

whenever $E^A \subset F_0^A$ and $E^B \subset F_0^B$. Really, this merely amounts to having the two random variables ω^A and ω^B be independent on their joint support $F_0^A \times F_0^B$.

5.2. Conditional Independence

Given two non-empty finite subsets F^A and F^B of the respective sample spaces Ω^A and Ω^B , say that the joint CCPS $P(\cdot|\cdot) \in \Delta_C(F^A \times F^B)$ is *conditionally independent* if there exist two component CCPS's $P^i(\cdot|\cdot) \in \Delta_C(F^i)$ ($i = A, B$) such that, whenever $\emptyset \neq E_1^i \subset E_2^i \subset F^i$ for $i = A, B$, then

$$P(E_1^A \times E_1^B | E_2^A \times E_2^B) = P^A(E_1^A | E_2^A) P^B(E_1^B | E_2^B). \quad (29)$$

A similar definition has been put forward independently by Battigalli (1991, 1992). Also, conditional independence is similar to BBD's (p. 74) concept of 'stochastically independent preferences.' In fact Battigalli and Veronesi (1992) have recently proved that, under suitable

decision-theoretic assumptions, preferences are stochastically independent if and only if subjective probabilities of states that are not Savage null are conditionally independent.

Note in particular how (29) implies that when $E^B = E_1^B = E_2^B$, then

$$P(E_1^A \times E^B | E_2^A \times E^B) = P^A(E_1^A | E_2^A). \quad (30)$$

So knowledge of the event E^B in the sample space Ω^B gives no information to affect the CCPS P^A on F^A . Nor does knowing the event $E^A \subset \Omega^A$ give any information to affect P^B on F^B , because

$$P(E^A \times E_1^B | E^A \times E_2^B) = P^B(E_1^B | E_2^B). \quad (31)$$

Conversely, the two conditions (30) and (31) jointly imply (29) because

$$P(E_1^A \times E_1^B | E_2^A \times E_2^B) = P(E_1^A \times E_1^B | E_2^A \times E_1^B) P(E_2^A \times E_1^B | E_2^A \times E_2^B). \quad (32)$$

It should be noted that conditional independence really is a strengthened form of almost sure independence. For suppose that the joint CCPS $P(\cdot|\cdot) \in \Delta_C(F^A \times F^B)$ is conditionally independent, and so satisfies (29) for a suitable pair of CCPS's P^i on F^i ($i = A, B$). Now let E^i ($i = A, B$) be the support of $P^i(\cdot|F^i)$. Then (29) implies that

$$P(E^A \times E^B | F^A \times F^B) = P^A(E^A | F^A) P^B(E^B | F^B) = 1 \quad (33)$$

and also

$$P(\{(\omega^A, \omega^B)\} | E^A \times E^B) = P^A(\{\omega^A\} | E^A) P^B(\{\omega^B\} | E^B) > 0 \quad (34)$$

for all pairs $(\omega^A, \omega^B) \in E^A \times E^B$, as required for almost sure independence of the joint CCPS on the support $E^A \times E^B$.

Almost sure independence, however, does not imply conditional independence. To see this, suppose that $F^i = \{\omega_j^i \mid j = 1, 2, 3\}$ for $i = A, B$. Consider the joint LCPS $\langle p_0, p_1 \rangle$ on the product space $F := F^A \times F^B$ which has

$$p_0(\omega_1^A, \omega_1^B) = p_0(\omega_1^A, \omega_2^B) = \frac{1}{3}; \quad p_0(\omega_2^A, \omega_1^B) = p_0(\omega_2^A, \omega_2^B) = \frac{1}{6}$$

on the support $F_0 = \{\omega_1^A, \omega_2^A\} \times \{\omega_1^B, \omega_2^B\}$, together with

$$p_1(\omega_3^A, \omega_1^B) = p_1(\omega_3^A, \omega_2^B) = p_1(\omega_1^A, \omega_3^B) = p_0(\omega_2^A, \omega_3^B) = \frac{1}{6}; \quad p_0(\omega_3^A, \omega_3^B) = \frac{1}{3}$$

on the support $F_1 = F \setminus F_0$. Then the first order joint distribution $p_0 \in \Delta(F^A \times F^B)$ is independent on its support F_0 because it corresponds to the product of the two distributions $p_0^A \in \Delta(F^A)$ and $p_0^B \in \Delta(F^B)$ given by

$$\begin{aligned} p_0^A(\omega_1^A) &= \frac{2}{3}; \quad p_0^A(\omega_2^A) = \frac{1}{3}; \quad p_0^A(\omega_3^A) = 0; \\ p_0^B(\omega_1^B) &= \frac{1}{2}; \quad p_0^B(\omega_2^B) = \frac{1}{2}; \quad p_0^B(\omega_3^B) = 0. \end{aligned} \quad (35)$$

Nevertheless, the three corresponding conditional LCPS's $\langle p_0^A(\cdot|\omega_j^B), p_1^A(\cdot|\omega_j^B) \rangle \in \Delta_L(F^A)$ ($j = 1, 2, 3$) must satisfy

$$p_0^A(\cdot|\omega_1^B) = p_0^A(\cdot|\omega_2^B) = p_0^A(\cdot); \quad p_0^A(\omega_1^A|\omega_3^B) = p_0^A(\omega_1^A|\omega_3^B) = \frac{1}{4}; \quad p_0^A(\omega_3^A|\omega_3^B) = \frac{1}{2}.$$

The corresponding conditional CCPS's are evidently not conditionally independent. Similarly, the three corresponding conditional LCPS's $\langle p_0^B(\cdot|\omega_j^A), p_1^B(\cdot|\omega_j^A) \rangle \in \Delta_L(F^B)$ ($j = 1, 2, 3$) must satisfy

$$p_0^B(\cdot|\omega_1^A) = p_0^B(\cdot|\omega_2^A) = p_0^B(\cdot); \quad p_0^B(\omega_1^B|\omega_3^A) = p_0^B(\omega_1^B|\omega_3^A) = \frac{1}{4}; \quad p_0^B(\omega_3^B|\omega_3^A) = \frac{1}{2}.$$

Once again, the corresponding conditional CCPS's are not conditionally independent.

5.3. Determining Joint Distributions

One defect of both LPS's and LCPS's is that in extensive games one often wants different players' strategies at different information sets to be stochastically independent. While the joint distribution of all the players' strategy choices can be specified as a LPS, this cannot easily be expressed in the usual way as the product of the LPS's attached to each individual player's strategy choice. Probabilities at different information sets need to be multiplied in order to compound them into the probabilities of different pure strategy profiles. With an entire lexicographic hierarchy of probabilities to keep track of, a suitable rule of multiplication is not immediately obvious.

In fact, a rather serious problem remains. For suppose that $F^A = \{\omega_1^A, \omega_2^A\}$ and $F^B = \{\omega_1^B, \omega_2^B\}$. Consider the two marginal CCPS's given by

$$P^A(\{\omega_1^A\}|F^A) = P^B(\{\omega_1^B\}|F^B) = 1; \quad P^A(\{\omega_2^A\}|F^A) = P^B(\{\omega_2^B\}|F^B) = 0. \quad (36)$$

Now we face the question of what the joint distribution on the space $F = F^A \times F^B$ must be. Unlike ordinary independent probabilities, however, knowledge of the independent marginal

CCPS's is insufficient to determine the joint distribution. Indeed, consider the following one dimensional continuum of joint LCPS's $\mathbf{p}^\alpha = \langle p_k^\alpha(\cdot) \rangle_{k=0}^{K^\alpha} \in \Delta_L(F)$ parametrized by α ($0 \leq \alpha \leq 1$). Suppose that $K^0 = K^1 = 3$, $K^\alpha = 2$ ($0 < \alpha < 1$), while

$$p_1^\alpha(\omega_1^A, \omega_2^B) = 1 - p_1^\alpha(\omega_2^A, \omega_1^B) = \alpha \quad (0 < \alpha < 1)$$

and that the hierarchical supports $\langle p_k^\alpha(\cdot) \rangle_{k=0}^{K^\alpha}$ satisfy

$$\begin{aligned} F_0^\alpha &= \{(\omega_1^A, \omega_1^B)\} \quad (0 \leq \alpha \leq 1); & F_1^\alpha &= \{(\omega_1^A, \omega_2^B), (\omega_2^A, \omega_1^B)\} \quad (0 < \alpha < 1) \\ F_1^0 &= F_2^1 = \{(\omega_1^A, \omega_2^B)\}; & F_1^1 &= F_2^0 = \{(\omega_2^A, \omega_1^B)\}; \\ F_2^\alpha &= F_3^0 = F_3^1 = \{(\omega_2^A, \omega_2^B)\} \quad (0 < \alpha < 1). \end{aligned}$$

All unspecified probabilities must then be either 0 or 1. As LCPS's, all these distributions are quite different, of course. Yet the corresponding conditional distributions are given by (36) and so all satisfy the independence conditions (29) above, as is easily verified. The problem is that our spaces of extended probabilities are still not rich enough to allow joint distributions to be inferred uniquely from their independent marginal distributions. Nor are the first two definitions of independence strict enough.

BBD (p. 74) suggest the obvious remedy, which is to consider products of (non-Archimedean) probability measures. Yet they do not consider in any detail which of the multitude of different possible non-Archimedean ordered fields should represent such measures (though the proof of their Theorem 6.1 relies on an ultrafilter construction due to Richter (1971), thus suggesting that they are willing to allow the full nonstandard space of hyperreals). The next section considers one very particular candidate for the field whose members should represent probabilities — namely, the smallest and simplest ordered field that has a chance of meeting all the requirements of decision and game theory because it is non-Archimedean and also extends the real line.

6. Non-Archimedean Probabilities

6.1. An Elementary Non-Archimedean Ordered Field

An (*algebraic*) *field* is a set G , together with the algebraic operations $+$ (addition), \cdot (multiplication), and the two corresponding identity elements 0 and 1. The set G must be closed under these two algebraic operations. The usual properties of real number arithmetic have to be satisfied — i.e., addition and multiplication both have to be commutative and associative, the distributive law must be satisfied, and every element of $x \in G$ must have both an additive inverse $-x$ and a multiplicative inverse $1/x$, except that $1/0$ is undefined. An *ordered field* also contains a binary relation $>$ which is a total order of G satisfying $1 > 0$, as well as the obvious properties that $x + y > x + z \implies y > z$ and that $x \cdot y > x \cdot z \implies y > z$ whenever $x > 0$. Equivalently, it should be true that $y > z \iff y - z > 0$, and that the set of positive elements in G is closed under addition and multiplication — as Robinson (1973), for instance, points out. Both the real line and the rationals are examples of ordered fields.

Robinson (1973, p. 88–9) also discusses the particular elementary ordered field which I shall denote by $\mathfrak{R}(\epsilon)$. It is the smallest field generated by combining the real line \mathfrak{R} with the single positive infinitesimal ϵ . It may help to think of ϵ as representing a fixed sequence of positive real numbers that converges to zero, such as $\epsilon = \langle 1/n \rangle_{n=1}^{\infty}$, or $\epsilon = \langle 10^{-n} \rangle_{n=1}^{\infty}$. Note first that, since $\mathfrak{R}(\epsilon)$ must be closed under addition and multiplication, its members have to be all the ‘rational’ functions which can be expressed as ratios

$$f(\epsilon) = \frac{A(\epsilon)}{B(\epsilon)} = \frac{a_0 + a_1 \epsilon + a_2 \epsilon^2 + \dots + a_n \epsilon^n}{b_0 + b_1 \epsilon + b_2 \epsilon^2 + \dots + b_m \epsilon^m} = \frac{\sum_{i=0}^n a_i \epsilon^i}{\sum_{i=0}^m b_i \epsilon^i} \quad (37)$$

of two polynomial functions $A(\epsilon), B(\epsilon)$ of the indeterminate ϵ with real coefficients; moreover not all the coefficients of the denominator $B(\epsilon)$ can be zero. Actually, after eliminating any leading zeros $a_0 = a_1 = \dots = a_{k-1} = b_0 = b_1 = \dots = b_{j-1} = 0$ and then dividing all coefficients of both the numerator and denominator of (37) by the leading non-zero coefficient b_j of the denominator, any member of $\mathfrak{R}(\epsilon)$ assumes the normalized form

$$f(\epsilon) = \frac{\sum_{i=k}^n a_i \epsilon^i}{\epsilon^j + \sum_{i=j+1}^m b_i \epsilon^i} \quad (38)$$

for some integers $j, k \geq 0$, where $a_k \neq 0$ unless $f(\epsilon) = 0$. Note too that each real number $r \in \mathfrak{R}$ can be expressed as $r = r/1$, and so has the form (38) with $j = k = m = n = 0$ and $a_0 = r$. Thus $\mathfrak{R} \subset \mathfrak{R}(\epsilon)$.

It remains to be shown that $\mathfrak{R}(\epsilon)$ really is an ordered field. The relation $>$ will be defined so that, when $f(\epsilon)$ is in the normalized form (38), then $f(\epsilon) > 0$ if and only if $a_k > 0$. This is entirely natural when ϵ is an infinitesimal, because this condition is equivalent to having the corresponding real valued rational function $f(x)$ be positive for all small positive real x . From this definition, it follows easily that either $f(\epsilon) > 0$ or $f(\epsilon) < 0$ unless $f(\epsilon) = 0$, so that the order is indeed total. And it is easy and routine to check that the corresponding set of positive elements is closed under addition and multiplication. In particular, $x - y > 0$ and $y - z > 0$ imply that $x - z = (x - y) + (y - z) > 0$, thus verifying that $>$ is transitive. Finally, $\mathfrak{R}(\epsilon)$ is non-Archimedean because $n\epsilon < 1$ for every integer n .

The non-Archimedean ordered field $\mathfrak{R}(\epsilon)$ is obviously much smaller and easier to describe than many others, including the hyperreal line ${}^*\mathfrak{R}$ which is generally used in non-standard analysis. It is also simpler than the field \mathcal{L} introduced by Levi-Civita (1892/3), whose members can be expressed as generalized power series $\sum_{k=1}^{\infty} a_k \epsilon^{\nu_k}$ with real coefficients a_k and real powers ν_k ($\nu = 1, 2, \dots$), such that the infinite sequence ν_k is strictly increasing and unbounded above. For other properties of the field \mathcal{L} , see Laugwitz (1968) and also Lightstone and Robinson (1975). For interesting historical and philosophical discussions of the infinitesimals, see Robinson (1966, ch. 10) as well as Stroyan and Luxemburg (1976).

This paper considers only probability distributions over finite sets. Probabilities will be represented by members of $\mathfrak{R}(\epsilon)$, satisfy finite additivity, and sum to 1 exactly. For countably additive non-Archimedean probability measures over general measurable spaces, however, it would be natural to extend $\mathfrak{R}(\epsilon)$ to a field that is closed under countable as well as finite summation. This suggests the need to consider the space $\mathfrak{R}^{\infty}(\epsilon)$ of normalized ratios of power series having the form

$$f^{\infty}(\epsilon) = \frac{\sum_{i=k}^{\infty} a_i \epsilon^i}{\epsilon^j + \sum_{i=j+1}^{\infty} b_i \epsilon^i}. \quad (39)$$

for some integers $j, k \geq 0$. For example, probabilities with values in $\mathfrak{R}^{\infty}(\epsilon)$ may prove important for the theory of games with compact strategy sets in arbitrary metric spaces — cf. Simon's (1987) discussion of 'local' trembling hand perfection.

Since each member of $\mathfrak{R}^{\infty}(\epsilon)$ can in fact be expressed as a single power series, it must be true that $\mathfrak{R}^{\infty}(\epsilon) \subset \mathcal{L}$. Thus $\mathfrak{R}(\epsilon) \subset \mathfrak{R}^{\infty}(\epsilon) \subset \mathcal{L} \subset {}^*\mathfrak{R}$.

6.2. Rational Probability Functions

At first Selten (1965, 1973) overcame the zero probability problem in games of perfect information by imposing ‘perfectness’ in proper subgames — i.e., subgames in which all players start with complete information. Later Selten (1975) introduced the concept of ‘trembling-hand perfect’ equilibria in which no player can ever be entirely sure what strategy another player is going to use. Instead there is always a small possibility of any strategy being played by mistake, no matter how bad the consequences of that strategy may be. A similar idea underlies Myerson’s (1978) concept of ‘proper equilibrium.’ It seems natural to represent ‘a small possibility’ by an infinitesimal probability, rather than by the kind of limiting small positive real probability which Selten and Myerson consider.

So suppose that, before normalization, the relative likelihoods $L(\omega; \epsilon)$ of different states $\omega \in F$ are represented by polynomials

$$L(\omega; \epsilon) := \ell_0(\omega) + \ell_1(\omega) \epsilon + \ell_2(\omega) \epsilon^2 + \cdots + \ell_K(\omega) \epsilon^K = \sum_{k=0}^K \ell_k(\omega) \epsilon^k \quad (40)$$

of degree K in the fixed positive infinitesimal ϵ . Here the coefficients $\ell_k(\omega)$ ($k = 0$ to K ; $\omega \in F$) are all assumed to be non-negative real numbers. Also each possible outcome ω in the finite set F is assumed to have a strictly positive likelihood. This is true, of course, if and only if $\sum_{k=0}^K \ell_k(\omega)$ is positive for every ω in F . When ϵ is regarded as a small positive number, such polynomial likelihood functions of degree one occur in Selten’s definition of trembling-hand perfect equilibrium. Polynomial likelihood functions of higher degree occur in Myerson’s definition of proper equilibrium. Note finally that we shall always assume that

$$\sum_{\omega \in F} \ell_0(\omega) > 0 \quad (41)$$

because there should be at least one $\omega \in F$ with a positive non-infinitesimal likelihood.

In order to transform them into rational probability functions, such polynomial likelihood functions must be normalized to make the sum over ω of the values of (40) identically equal to one. Thus, (40) needs to be divided by the obvious normalizing factor

$$L(\epsilon) := \sum_{\omega \in F} L(\omega; \epsilon) = \ell_0 + \ell_1 \epsilon + \ell_2 \epsilon^2 + \cdots + \ell_K \epsilon^K = \sum_{k=0}^K \ell_k \epsilon^k \quad (42)$$

where $\ell_k := \sum_{\omega \in F} \ell_k(\omega)$ for $k = 0$ to K , and so $\ell_0 > 0$ because of our assumption (41). The result of this normalization will be a strictly positive *rational probability function* (or

RPF) on F . This takes the form

$$f(\omega; \epsilon) := \frac{L(\omega; \epsilon)}{L(\epsilon)} = \frac{\ell_0(\omega) + \ell_1(\omega)\epsilon + \ell_2(\omega)\epsilon^2 + \cdots + \ell_K(\omega)\epsilon^K}{\ell_0 + \ell_1\epsilon + \ell_2\epsilon^2 + \cdots + \ell_K\epsilon^K} = \frac{\sum_{k=0}^K \ell_k(\omega)\epsilon^k}{\sum_{k=0}^K \ell_k\epsilon^k} \quad (43)$$

and so specifies a positive non-Archimedean probability $f(\omega; \epsilon) \in \mathfrak{R}(\epsilon)$ for each $\omega \in F$. It does no harm to normalize (43) further by dividing every coefficient in either the numerator or the denominator by the positive number ℓ_0 . The effect will be a *normalized* RPF taking the form

$$f(\omega; \epsilon) := \frac{\ell_0(\omega) + \ell_1(\omega)\epsilon + \ell_2(\omega)\epsilon^2 + \cdots + \ell_K(\omega)\epsilon^K}{1 + \ell_1\epsilon + \ell_2\epsilon^2 + \cdots + \ell_K\epsilon^K} = \frac{\sum_{k=0}^K \ell_k(\omega)\epsilon^k}{1 + \sum_{k=1}^K \ell_k\epsilon^k} \quad (44)$$

for suitably redefined constants $\ell_k(\omega)$ ($\omega \in F$) and ℓ_k ($k = 0$ to K) that still satisfy the requirements that $0 \leq \ell_k(\omega)$ and $\sum_{\omega \in F} \ell_k(\omega) = \ell_k$ ($k = 0$ to K), where $\ell_0 = 1$. Thus each $f(\omega; \epsilon)$ has been expressed in the normalized form (38). Note that $\ell_0(\omega)$ ($\omega \in F$) is an ordinary probability distribution in $\Delta(F)$. Moreover $\sum_{k=0}^K \ell_k(\omega) > 0$ for each $\omega \in F$.

Let $\Delta^0(F; \epsilon)$ denote the set of all such RPF's on the finite set F . Note that neither $(\frac{1}{2}, \frac{1}{2}, \epsilon)$ nor $(1 - \epsilon, \epsilon)$ are possible values of RPF's. The first needs normalizing to become $(1 + \epsilon)^{-1}(\frac{1}{2}, \frac{1}{2}, \epsilon)$. The second is excluded because it has a negative coefficient; a similar but different effect is produced by the valid RPF with values $(1 + \epsilon)^{-1}(1, \epsilon)$.

Of course, each distribution $p(\cdot) \in \Delta^0(F)$ can be identified with the particular RPF of the form (44) above, with $K = 0$ and $\ell_0(\omega) = p(\omega) > 0$ for all $\omega \in F$. Thus $\Delta^0(F; \epsilon)$ is an extension of the set $\Delta^0(F)$ of interior probability distributions on F .

Now, one might be tempted to eliminate redundancy among the different terms of the likelihood polynomial (44) by requiring every ℓ_k in the denominator to be positive. Of course, the set of all such restricted RPF's is still closed under addition and multiplication. However, it is *not* closed under division, as is required for all conditional probabilities to be in the same space. Indeed, suppose that $F = \{\omega_j \mid j = 0, 1, 2\}$ and that $f(\omega; \epsilon) \in \Delta^0(F; \epsilon)$ is given by

$$f(\omega_j; \epsilon) = \frac{\epsilon^j}{1 + \epsilon + \epsilon^2} \quad (45)$$

for $j = 0, 1, 2$. Then the common denominator of the corresponding non-Archimedean conditional probabilities

$$f(\omega_0 \mid \{\omega_0, \omega_2\}; \epsilon) = \frac{1}{1 + \epsilon^2}; \quad f(\omega_2 \mid \{\omega_0, \omega_2\}; \epsilon) = \frac{\epsilon^2}{1 + \epsilon^2} \quad (46)$$

has terms in ϵ^2 , but an important and crucial absence of any term in ϵ . For this reason, it is important to consider the whole space $\Delta^0(F; \epsilon)$ of all possible RPF's, without any further restrictions. This will be confirmed in later work that considers consequentialist behaviour norms in decision trees having RPF's at their chance nodes.

6.3. Lexicographic Rational Probability Functions

In order to have them correspond to LPS's, rational probability function have to be normalized once more. This second normalization requires making the different coefficients $\ell_k(\omega)$ ($\omega \in F$) of each power ϵ^k ($k = 0$ to K) in the numerator of ϵ in (43) into the probability distribution $p_k \in \Delta(F)$ over all the possible values of ω defined by

$$p_k(\omega) = \ell_k(\omega)/\ell_k. \quad (47)$$

The coefficient of ϵ^k in the denominator becomes 1. This is only possible, of course, in case $\ell_k > 0$ ($k = 0$ to K), and so we assume that this is true. The result of this second normalization will be a *lexicographic* RPF of the form

$$f(\omega; \epsilon) := \frac{p_0(\omega) + p_1(\omega)\epsilon + p_2(\omega)\epsilon^2 + \cdots + p_K(\omega)\epsilon^K}{1 + \epsilon + \cdots + \epsilon^K} = \frac{\sum_{k=0}^K p_k(\omega)\epsilon^k}{\sum_{k=0}^K \epsilon^k}. \quad (48)$$

Let $\Delta_R^*(F)$ denote the set of all such lexicographic RPF's. The normalized coefficients which appear in the numerator of (48) clearly correspond uniquely to the LPS $\langle p_k \rangle_{k=0}^K \in \Delta_L^*(F)$. So there is an obvious one-to-one correspondence ψ_{LR} between the space $\Delta_L^*(F)$ of LPS's and the space $\Delta_R^*(F)$. Since $\Delta_L^*(F)$ is too restrictive for our theory, so is $\Delta_R^*(F)$.

6.4. Conditional Rational Probability Functions

When the same one-to-one correspondence $\psi_{LR} : \Delta_L^*(F) \rightarrow \Delta_R^*(F)$ is restricted to the domain $\Delta_L(F)$ of LCPS's, it has a range that will be denoted by

$$\Delta_R(F) := \psi_{LR}(\Delta_L(F)) \subset \Delta_R^*(F). \quad (49)$$

Then it is obvious that the members of $\Delta_R(F)$ are lexicographic RPF's meeting the extra requirement that the probability distributions $p_k(\cdot) \in \Delta(F)$ ($k = 0$ to K) whose values appear in the numerator of (48) have disjoint supports, which will be denoted by F_k ($k = 0$ to K). In fact, given any LCPS $\langle p_k \rangle_{k=0}^K \in \Delta_L(F)$, for each $\omega \in F$ there must be a unique

integer $k(\omega)$ in the range $k = 0$ to K such that $\omega \in F_{k(\omega)}$ and so $p_{k(\omega)}(\omega) > 0$. Thus, for each $\omega \in F$, the numerator of (48) must have only one term $p_{k(\omega)}(\omega) \epsilon^{k(\omega)}$ with a positive coefficient. Therefore each member of $\Delta_R(F)$ can be regarded as a unique *conditional* RPF of the simple form

$$f(\omega; \epsilon) = \frac{p_{k(\omega)}(\omega) \epsilon^{k(\omega)}}{1 + \epsilon + \epsilon^2 + \dots + \epsilon^K}. \quad (50)$$

Of course, all the powers of ϵ between 0 and K must occur in the numerator of (50) for some $\omega \in F$. So the maximum possible power ϵ^K must satisfy $K + 1 \leq \#F$, the size of F . Also the mapping ψ_{LR} restricted to $\Delta_L(F)$ is obviously a one-to-one correspondence between the two sets $\Delta_L(F)$ and $\Delta_R(F)$. Now, Section 4 showed how $\Delta_L(F)$ is equivalent to both the spaces $\Delta_C(F)$ and $\Delta_M(F)$ of ‘canonical form’ extended probabilities. So therefore is $\Delta_R(F)$. That is why its members are called ‘conditional’ RPF’s.

6.5. Multiplicative Independence

Let F^A and F^B be two non-empty finite subsets of the respective sample spaces Ω^A and Ω^B . Now that probabilities have been given values in a non-Archimedean field, the joint RPF $f(\omega^A, \omega^B; \epsilon) \in \Delta^0(F^A \times F^B; \epsilon)$ can be defined as *multiplicatively independent* if there exist RPF’s $f^A(\omega^A; \epsilon) \in \Delta^0(F^A; \epsilon)$ and $f^B(\omega^B; \epsilon) \in \Delta^0(F^B; \epsilon)$ for which

$$f(\omega^A, \omega^B; \epsilon) \equiv f^A(\omega^A; \epsilon) f^B(\omega^B; \epsilon). \quad (51)$$

Of course, this is a natural extension from $\Delta(F^A \times F^B)$ to $\Delta^0(F^A \times F^B; \epsilon)$ of the usual definition of independence. And it is immediately obvious that it strengthens the definition of conditional independence given in Section 5.2. It is the strongest of the three definitions of independence given by BBD (p. 74). Finally, given two multiplicatively independent RPF’s, the corresponding joint RPF is simply found by multiplication. This is the definition of independence which I shall use in later work.

7. Summary and Conclusion

This paper commenced with two examples showing the inadequacy of ordinary probabilities for the purposes of orthodox decision and game theory. Thereafter it considered extended probabilities in the following three spaces:

- (a) $\Delta_C(F)$ of complete conditional probability systems (or CCPS's), as defined in Section 3.2 (cf. Myerson, 1986);
- (b) $\Delta_L(F)$ of lexicographic conditional probability systems (or LCPS's), restricted so that the distributions in each hierarchy have disjoint supports, as defined in Section 3.3 (cf. Blume, Brandenburger and Dekel, 1991a, b);
- (c) $\Delta_M(F)$ of 'consistent conditional systems' of logarithmic likelihood ratio functions (or LLRF's), as defined in Section 3.4 (cf. McLennan, 1989a, b).

It was shown in Section 4 that all these three spaces are actually equivalent, and can also be given metrics which make them homeomorphic and compact. Thus they are really a single space of 'canonical' extended probabilities.

Section 3.3 also introduced the space:

- (d) $\Delta_L^*(F)$ of unrestricted lexicographic probability systems (or LPS's), whose supports may overlap.

This is somewhat more extensive than the canonical space described by (a)–(c) above.

When there are several random variables, Blume, Brandenburger and Dekel (1991a) proposed three different definitions of independence. These were characterized as:

- (i) almost sure independence, which is the weakest (Section 5.1);
- (ii) conditional independence, which seems the most natural for complete conditional probability systems, at least (Section 5.2);
- (iii) multiplicative independence, which is the strongest (Section 6.5), but is not well defined in the space of canonical extended probabilities, nor in the space $\Delta_L^*(F)$ of general unrestricted LPS's.

It turns out that the difficulty with the third definition of independence is crucial. For an important property of standard probabilities is that, once independent marginal distributions are known, so is the joint distribution. In decision trees, for instance, the joint distribution

of random moves at all chance nodes of the tree can be inferred from the distributions of independent random moves at each separate chance node. For extended probabilities, the first two notions of independence set out above fail this test.

In order to overcome this deficiency, Section 6 introduced the elementary non-Archimedean ordered field $\mathfrak{R}(\epsilon)$ generated by the real line \mathfrak{R} together with the single positive infinitesimal ϵ . Its members are values of rational functions evaluated at ϵ . This suggests considering rational probability functions — i.e., probabilities represented by rational functions in $\mathfrak{R}(\epsilon)$. Because $\mathfrak{R}(\epsilon)$ is much simpler than the space ${}^*\mathfrak{R}$ of hyperreals, such probabilities are much simpler than those considered by Narens (1974b, 1985 pp. 282–293) or Skala (1975, p. 119).

Finally, Section 6 introduced three new spaces of probabilities with values in $\mathfrak{R}(\epsilon)$:

- (e) $\Delta^0(F; \epsilon)$, consisting of general rational probability functions (or RPF's), as defined in Section 6.2;
- (f) $\Delta_R^*(F)$, consisting of lexicographic RPF's, as defined in Section 6.3;
- (g) $\Delta_R(F)$, consisting of conditional RPF's, as defined in Section 6.4.

In fact $\Delta_R(F)$ was constructed so as to be equivalent to each of the first three equivalent spaces of canonical extended probabilities that were described in (a)–(c) above. Obviously $\Delta_R(F)$ is a proper subset of $\Delta_R^*(F)$, which was constructed to be equivalent to the space $\Delta_L^*(F)$ of unrestricted LPS's. But $\Delta_R^*(F)$ in turn is a proper subset of $\Delta^0(F; \epsilon)$, which is the only one of these spaces that is rich enough for the crucial property of multiplicative independence to make sense. For this reason, none of the first four spaces (a)–(d) of extended probabilities is adequate. Nor are the spaces (f) and (g). Instead, only the complete space (e), or some other space of even further enriched non-Archimedean measures of probability, will work. That is the main conclusion of the paper.

The basic theory of 'consequentialist' behaviour in decision trees with both random moves and uncertain states of the world was presented in Hammond (1988b). A natural extension of that theory, to be explored in Hammond (1992) and later work, attaches non-Archimedean probabilities to random moves at each chance node of any finite decision tree. Thereafter will come the extension of consequentialist decision theory to games. This will prove the crucial test of whether the non-Archimedean probabilities proposed here really are sufficient to overcome the zero probability problem which has motivated this paper.

Acknowledgements

In the first place, Patrick Suppes has offered much valuable intellectual and moral support over the years for all my research. For this paper in particular, he was the first to alert me to Rényi's valuable contribution, and has since made a number of insightful comments.

During the academic year 1986–7, some initial work for this and related papers was generously supported by the Guggenheim Foundation and by C.O.R.E. at the Université Catholique de Louvain. Much of a preliminary draft was written during a fruitful visit to the Department of Economics at Princeton University in the Spring of 1987. Later, in Spring 1988 I received support from the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 303 while visiting the University of Bonn.

The work owes its origins to particularly useful interaction with Adam Brandenburger and Eddie Dekel. C.O.R.E. also provided the opportunity for many helpful discussions with Jean-François Mertens. Earlier versions were incorporated in presentations to the conference on 'Sequential Equilibria and Structural Stability' at Luminy, January 1987, to the workshop on 'Game Theory in the Behavioral Sciences' at the Zentrum für Interdisziplinäre Forschung at the University of Bielefeld in April 1988, and to the mathematical economics seminars at Princeton, Stanford, Harvard, and Bonn Universities. Donald Brown has offered me invaluable expert advice concerning non-standard analysis. Other improvements and encouragement have resulted from helpful comments by Bill Zame, Amos Tversky, Jeroen Swinkels and several seminar participants, Reinhard Selten, Barry Nalebuff, William Harper, Avinash Dixit, Gabrielle Demange and Andrew Caplin in particular. My thanks to all the above individuals and institutions, without absolving myself of full responsibility for any remaining errors and deficiencies.

REFERENCES

- Barnard, G.A.: 1949, 'Statistical Inference', *Journal of the Royal Statistical Society, Series B* **11**, 115–139.
- Battigalli, P.: 1991, 'Strategic Independence, Generally Reasonable Extended Assessments, and Consistent Assessments', preprint, Istituto di Economia Politica, Università 'L. Bocconi', Milan.

- Battigalli, P.: 1992, 'Strategic Rationality Orderings', preprint, Istituto di Economia Politica, Università 'L. Bocconi', Milan.
- Battigalli, P. and P. Veronesi: 1992, 'A Note on Stochastic Independence without Savage-Null Events', preprint, Istituto di Economia Politica, Università 'L. Bocconi', Milan.
- Blackwell, D. and L.E. Dubins: 1975, 'On Existence and Non-existence of Proper, Regular, Conditional Distributions', *The Annals of Probability* **3**, 741–752.
- Blackwell, D. and C. Ryll-Nardzewski: 1963, 'Non-existence of Everywhere Proper Conditional Distributions', *Annals of Mathematical Statistics* **34**, 223–225.
- Blume, L., A. Brandenburger and E. Dekel: 1991a, 'Lexicographic Probabilities and Choice Under Uncertainty', *Econometrica* **59**, 61–79.
- Blume, L., A. Brandenburger and E. Dekel: 1991b, 'Lexicographic Probabilities and Equilibrium Refinements', *Econometrica* **59**, 81–98.
- Chernoff, H.: 1954, 'Rational Selection of Decision Functions', *Econometrica* **22**, 422–443.
- Chipman, J.S.: 1960, 'Foundations of Utility', *Econometrica* **28**, 193–224.
- Chipman, J.S.: 1971a, 'On the Lexicographic Representation of Preference Orderings', in Chipman et al., ch. 14, pp. 276–288.
- Chipman, J.S.: 1971b, 'Non-Archimedean Behavior under Risk: An Elementary Analysis — With Application to the Theory of Assets', in Chipman et al., ch. 15, pp. 289–318.
- Chipman, J.S., L. Hurwicz, M.K. Richter and H. Sonnenschein (eds.): 1971, *Preferences, Utility and Demand: A Minnesota Symposium*, Harcourt Brace Jovanovic, New York.
- Császár, Á.: 1955, 'Sur la structure des espaces de probabilité conditionnelle', *Acta Mathematica Academiae Scientiarum Hungaricae* **6**, 337–361.
- Davidson, D., J.C.C. McKinsey and P. Suppes: 1955, 'Outlines of a Formal Theory of Value, I', *Philosophy of Science* **22**, 140–160.
- Davidson, D. and P. Suppes: 1956, 'A Finitistic Axiomatization of Subjective Probability and Utility', *Econometrica* **24**, 264–275.

- Davidson, D., P. Suppes and S. Siegel: 1957, *Decision Making: An Experimental Approach*, Stanford University Press, Stanford.
- de Finetti, B.: 1936, 'Les probabilités nulles', *Bulletin des Sciences Mathématiques* **60** (première partie), 275–288.
- de Finetti, B.: 1949, 1972, 'Sull'impostazione assiomatica del calcolo della probabilità', *Annali Triestini dell'Università di Trieste* **19**, 29–81; translated as 'On the Axiomatization of Probability Theory' in *Probability, Induction and Statistics: The art of guessing*, John Wiley, New York, ch. 5, pp. 67–113.
- Fudenberg, D. and J. Tirole: 1991, 'Perfect Bayesian Equilibrium and Sequential Equilibrium', *Journal of Economic Theory* **53**, 236–260.
- Good, I.J.: 1950, *Probability and the Weighing of Evidence*, Charles Griffin & Co, London.
- Hammond, P.J.: 1988a, 'Consequentialism and the Independence Axiom', in B. Munier (ed.), *Risk, Decision, and Rationality: Proceedings of the 3rd International Conference on the Foundations and Applications of Utility, Risk, and Decision Theories*, D. Reidel, Dordrecht, pp. 503–516.
- Hammond, P.J.: 1988b, 'Consequentialist Foundations for Expected Utility', *Theory and Decision* **25**, 25–78.
- Hammond, P.J.: 1992, 'Consequentialism, Non-Archimedean Probabilities, and Lexicographic Expected Utility', preprint, Department of Economics, Stanford University.
- Harper, W.L.: 1975, 'Rational Belief Change, Popper Functions, and Counterfactuals', *Synthese* **30**, 221–262.
- Harper, W.L., R. Stalnaker and G. Pearce (eds.): 1981, *IFS: Conditionals, Beliefs, Decisions, Chance and Time*, D. Reidel, Dordrecht.
- Hausner, M.: 1954, 'Multidimensional Utilities', in Thrall et al., ch. 12, pp. 167–180.
- Keynes, J.M.: 1921, *A Treatise on Probability Theory*, Macmillan, London.
- Kolmogorov, A.N.: 1933, 1956, *Grundbegriffe der Wahrscheinlichkeitsrechnung*; translated as *Foundations of the Theory of Probability*, Springer, Berlin, and Chelsea, New York.

- Koopman, B.O.: 1940, ‘The Axioms and Algebra of Intuitive Probability’, *Annals of Mathematics* **Ser. 2, 41**, 269–292.
- Krantz, D.H., R.D. Luce, P. Suppes and A. Tversky: 1971, *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*, Academic Press, New York.
- Krantz, D.H., R.D. Luce, P. Suppes and A. Tversky: 1990, *Foundations of Measurement, Vol. III: Representation, Axiomatization and Invariance*, Academic Press, New York.
- Kreps, D. and R. Wilson: 1982, ‘Sequential Equilibrium’, *Econometrica* **50**, 863–894.
- Laugwitz, D.: 1968, ‘Eine nichtarchimedische Erweiterung angeordneter Körper’, *Mathematische Nachrichten* **37**, 225–236.
- Levi, I.: 1980, *The Enterprise of Knowledge*, M.I.T. Press, Cambridge, Mass.
- Levi-Civita, T.: 1892/3, ‘Sugli infiniti ed infinitesimali attuali: quali elementi analitici’, *Atti Istituto Veneto di scienze, lettere, ed arti* **7**, 1765–1815; reprinted in *Opere Matematiche di Tullio Levi-Civita: Memorie e Note: Volume primo (1893–1900)*, (Nicola Zanichelli, Bologna, 1954), pp. 1–39.
- Lewis, D.K.: 1973, *Counterfactuals*, Harvard University Press, Cambridge, Mass.
- Lightstone, A.H. and A. Robinson: 1975, *Nonarchimedean Fields and Asymptotic Expansions*, North-Holland, Amsterdam.
- Lindley, D.V.: 1965, *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 1: Probability*, Cambridge University Press, Cambridge.
- Luce, R.D. and P. Suppes: 1965, ‘Preferences, Utility, and Subjective Probability’, in R.D. Luce et al. (ed.), *Handbook of Mathematical Psychology, III*, John Wiley, New York.
- Marschak, J. and R. Radner: 1972, *Economic Theory of Teams*, Yale University Press, New Haven.
- McLennan, A.: 1989a, ‘The Space of Conditional Systems is a Ball’, *International Journal of Game Theory* **18**, 125–139.

- McLennan, A.: 1989b, ‘Consistent Conditional Systems in Noncooperative Game Theory’, *International Journal of Game Theory* **18**, 141–174.
- Myerson, R.: 1978, ‘Refinements of the Nash Equilibrium Concept’, *International Journal of Game Theory* **7**, 73–80.
- Myerson, R.: 1986, ‘Multistage Games with Communication’, *Econometrica* **54**, 323–358.
- Narens, L.: 1974a, ‘Measurement without Archimedean Axioms’, *Philosophy of Science* **41**, 374–393.
- Narens, L.: 1974b, ‘Minimal Conditions for Additive Conjoint Measurement and Qualitative Probability’, *Journal of Mathematical Psychology* **11**, 404–430.
- Narens, L.: 1985, *Abstract Measurement Theory*, MIT Press, Cambridge, Mass.
- Popper, K.R.: 1934, 1959, *Logik der Forschung*; translated as *The Logic of Scientific Discovery*, Vienna, Springer Verlag, and Basic Books, New York.
- Popper, K.R.: 1938, ‘A Set of Independent Axioms for Probability’, *Mind* **47**, 275ff.
- Rényi, A.: 1955, ‘On a New Axiomatic Theory of Probability’, *Acta Mathematica Academiae Scientiarum Hungaricae* **6**, 285–335.
- Rényi, A.: 1956, ‘On Conditional Probability Spaces Generated by a Dimensionally Ordered Set of Measures’, *Theory of Probability and its Applications* **1**, 61–71; reprinted as Paper 120, pp. 547–557 of *Selected Papers of Alfréd Rényi, I: 1948–1956*, Akadémia Kiadó, Budapest, 1976.
- Rényi, A.: 1964, ‘Sur les espaces simples de probabilités conditionnelles’, *Annales Institut H. Poincaré*, Nouvelle série, Section B **1**, 3–21; reprinted as Paper 237, pp. 284–302 of *Selected Papers of Alfréd Rényi, III: 1962–1970*, Akadémia Kiadó, Budapest, 1976.
- Rényi, A.: 1970, *Probability Theory*, Elsevier, New York.
- Richter, M.K.: 1971, ‘Rational Choice’, in Chipman et al., ch. 2, pp. 29–58.
- Robinson, A.: 1951, *On the Metamathematics of Algebra*, North-Holland, Amsterdam.
- Robinson, A.: 1966, *Non-Standard Analysis*, North-Holland, Amsterdam.

- Robinson, A.: 1973, 'Function Theory on Some Nonarchimedean Fields', *American Mathematical Monthly: Papers in the Foundations of Mathematics* **80**, S87–S109.
- Selten, R.: 1965, 'Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragerträge', *Zeitschrift für die gesamte Staatswissenschaft* **121**, 301–324 and 667–689.
- Selten, R.: 1973, 'A Simple Model of Imperfect Competition, where 4 Are Few and 6 Are Many', *International Journal of Game Theory* **2**, 141–201.
- Selten, R.: 1975, 'Re-examination of the Perfectness Concept for Equilibrium Points of Extensive Games', *International Journal of Game Theory* **4**, 25–55.
- Selten, R. and U. Leopold: 1982, 'Subjunctive Conditionals in Decision Theory and Game Theory', in W. Stegmüller, W. Balzer and W. Spohn (ed.), *Philosophy in Economics: Proceedings, Munich, July 1981*, Springer Verlag, Berlin, pp. 191–200.
- Simon, L.: 1987, 'Local Perfection', *Journal of Economic Theory* **43**, 134–156.
- Skala, H.J.: 1975, *Non-Archimedean Utility Theory*, D. Reidel, Dordrecht.
- Spohn, W.: 1988, 'Ordinal Conditional Functions: A Dynamic Theory of Epistemic States', in W.L. Harper and B. Skyrms (ed.), *Causation in Decision, Belief Change, and Statistics*, Kluwer, Dordrecht, pp. 105–134.
- Spohn, W.: 1990, 'A General Non-Probabilistic Theory of Inductive Reasoning', in R.D. Schachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer (ed.), *Uncertainty in Artificial Intelligence 4*, North-Holland, Amsterdam, pp. 149–158.
- Stalnaker, R.: 1970, 'Probability and Conditionals', *Philosophy of Science* **37**, 64–80; reprinted in Harper et al., pp. 107–128.
- Stroyan, P. and W.A.J. Luxemburg: 1976, *Introduction to the Theory of Infinitesimals*, Academic Press, New York.
- Suppes, P.: 1956, 'The Role of Subjective Probability and Utility in Decision Making', in J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5*, University of California Press, Berkeley, pp. 61–73.
- Suppes, P.: 1961, 'Behavioristic Foundations of Utility', *Econometrica* **29**, 186–202.

Suppes, P.: 1969, *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, D. Reidel, Dordrecht.

Suppes, P.: 1972, *Axiomatic Set Theory* (2nd edn.), Dover, New York.

Suppes, P.: 1984, *Probabilistic Metaphysics*, Basil Blackwell, Oxford.

Thrall, R.M.: 1954, 'Applications of Multidimensional Utility Theory', in Thrall et al., ch. 13, pp. 181–186.

Thrall, R.M., C.H. Coombs and R.L. Davis (eds.): 1954, *Decision Processes*, John Wiley, New York.

Peter J. Hammond

Department of Economics

Stanford University, CA 94305-6072, U.S.A.