Consequentialism, Non-Archimedean Probabilities, and Lexicographic Expected Utility

Peter J. Hammond

Department of Economics, Stanford University, CA 94305-6072, U.S.A.

Abstract

Earlier work (Hammond, 1988a, b) on dynamically consistent "consequentialist" behaviour in decision trees was unable to treat zero probability events satisfactorily. Here the rational probability functions considered in Hammond (1994), as well as other non-Archimedean probabilities, are incorporated into decision trees. As before, the consequentialist axioms imply the existence of a preference ordering satisfying independence. In the case of rational probability functions, those axioms, together with continuity and a new refinement assumption, imply the maximization of a somewhat novel lexicographic expected utility preference relation. This is equivalent to maximization of expected utility in the ordering of the relevant non-Archimedean field.

Non-Archimedean Expected Utility

ma la natura la dà sempre scema
similemente operando all'artista
c'ha l'abito dell'arte e man che trema

Dante, *La Divina Commedia, Vol. III: Paradiso* (Canto XIII, 76–78)[†]

1. Introduction and Outline

1.1. Consequentialism in Single Person Decision Theory

The consequentialist approach to single-person decision theory, with uncertainty described by specified objective probabilities, was previously described in Hammond (1988a, b). In fact the latter paper also deals with states of nature and subjective probabilities. Here non-Archimedean subjective probabilities will not be considered at all, but left for later work.

This consequentialist approach is based on three axioms. Of these, the first requires behaviour to be well defined for an (almost) unrestricted domain of finite decision trees with: (i) decision nodes where the agent makes a move; (ii) chance nodes at which random moves occur with specified positive probabilities; (iii) and terminal nodes which result in a single consequence within a specified domain of relevant consequences. It follows that behaviour in any such tree results in a set of probability distributions over consequences. The second axiom requires behaviour to be dynamically consistent in continuation subtrees, in the sense that behaviour at any decision node is the same in a subtree as in a full tree. And the third is the "consequentialist choice" axiom, requiring that the set of possible random consequences of behaviour in any decision tree of the domain be explicable as the choice of desirable random consequences from the set of random consequences that the tree makes feasible.

[†] "But Nature fumbles, with no sure command
Over her tools, like an artificer
Who know his trade but has a trembling hand."

Dante Alighieri the Florentine, *The Divine Comedy, Cantica III: Paradise*; translated (rather freely) by D.L. Sayers and B. Reynolds (Penguin: Harmondsworth, 1962).

The earlier work then showed how these three consequentialist axioms imply the existence of a preference ordering (i.e., a complete and transitive weak preference relation) defined on the space of random consequences. Moreover, this ordering must satisfy Samuelson's (1952) independence axiom. In this way, two of the most important and even controversial axioms of standard decision theory become implications of apparently weaker and possibly more appealing axioms of the consequentialist reformulation. Thereafter, an extra condition of continuity of behaviour with respect to changing probabilities leads to the preference ordering having an expected utility representation.

1.2. Consequentialism in Game Theory

This paper is the second in a series whose purpose is to extend the scope of this consequentialist approach from single-person decision trees to multi-person extensive games. In fact, a decision tree is nothing more than a one person "consequentialist extensive game" of perfect and complete information. The difference from the usual notion of a game in extensive form comes about because no payoffs are specified. Instead, terminal nodes are assumed to result in pure consequences within the specified domain of consequences. Then the existence of a payoff function defined on this domain of consequences is not assumed, but becomes an important implication of consequentialism.

The main obstacle to this extension of consequentialism comes about because the earlier single person theory excludes zero probability chance moves — it is in this sense that the domain of allowable decision trees is *almost* rather than fully unrestricted. Such a restriction is excusable in single person decision theory where there are no good reasons for retaining zero probability events. In multi-person game theory, however, testing to see whether a particular profile of strategies for each player constitutes an equilibrium involves seeing what happens when any player deviates and then all the other players react according to their presumed equilibrium strategies. Yet these reactions are behaviour in the face of events which are supposed to have probability zero, since that is supposed to be the probability of any player deviating from equilibrium. So one is faced with the need to update probabilities in a Bayesian manner even though a zero probability event has occurred. For this reason, the zero probability restriction makes it difficult to apply consequentialist decision theory to multi-person games in a way that yields subgame perfect or other kinds of refined equilibria. The task of this work is to remove this burdensome restriction.

Not surprisingly, the zero probability problem has already led several game theorists to extend the space of ordinary probabilities in various ways. In particular, Selten (1975) and Myerson (1978) considered "trembles." Kreps and Wilson (1982), and then Blume, Brandenburger and Dekel (1991a, b) considered lexicographic hierarchies of probabilities. Myerson (1986) considered complete conditional probability systems. Finally, McLennan (1989a, b) considered logarithmic likelihood ratio functions whose values can be infinite. Hammond (1994) shows that all of these different extensions, when suitably formulated, are in fact equivalent to each other, and also to a particular space of "conditional rational probability functions."

In decision trees, it is usual to specify independent probabilities at each separate chance node. After all, in principle any causes of dependence can and should be modelled within the structure of the tree itself. Then, however, it is particularly important to be sure that the entire joint distribution of all chance moves is determined uniquely from the marginal probability distributions over the chance moves at each separate chance node of the decision tree. In the case of ordinary probabilities, this is trivial because joint probabilities are found by simply multiplying appropriate marginal probabilities. Yet Hammond (1994) also shows how the equivalent sets of extended probabilities mentioned in the previous paragraph all fail this crucial test. That is, many different joint extended probability distributions over nature's possible strategies in the decision tree can arise from the same collection of independent marginal probability distributions over nature's moves at different chance nodes. This is because such extended probabilities lack the structure of an algebraic field in which the operation of multiplication is well defined.

1.3. Consequentialism with Non-Archimedean Probabilities

So it seems necessary to work with a richer space of extended probabilities, for which multiplication and other algebraic field operations are well defined. There should also be an ordering relation rich enough to give meaning to the statements that probabilities are non-negative or positive and that they are larger or smaller. Now, within an ordered algebraic field such as the real line $\Re$, the *Archimedean axiom* states that, for any positive number r, no matter how small, there is an integer n for which $n\,r > 1$. Overcoming the zero probability problem seems to require some kind of *non-Archimedean* ordered field with some positive elements so small that this axiom is violated. So any such field must have at least one positive "infinitesimal" element ϵ with the property that $n\,\epsilon < 1$ for every positive integer n, even though ϵ is positive. Of course any such positive infinitesimal must be smaller than any positive real number; in particular, it cannot be a real number itself.

Section 2 below therefore begins by briefly reviewing the definition and key properties of non-Archimedean ordered fields. Particular attention is paid to the elementary field $\Re(\epsilon)$ whose use was recommended in Hammond (1994). All its members are rational functions of a single indeterminate infinitesimal denoted by ϵ. For any finite support in an appropriate sample space, Section 2 proceeds to consider a corresponding set of elementary non-Archimedean probabilities. In the case when the field is $\Re(\epsilon)$, such probabilities are called "rational probability functions" (or RPFs).

It then becomes natural to consider in Section 3 decision trees having RPFs or other non-Archimedean instead of ordinary positive probabilities attached to each chance move in the tree. Behaviour in such trees gives rise to *non-Archimedean consequences* — i.e., RPFs or more general non-Archimedean conditional probability distributions over the domain of relevant consequences. Within the corresponding domain of finite *non-Archimedean consequentialist decision trees* the previous consequentialist axioms can be applied almost without change. Most implications are also the same as in Hammond (1988a, b), since virtually all the arguments in those two papers apply to probabilities taking values in a general ordered field, and not just to those taking values in the Archimedean ordered field $\Re$. In particular, the three "consequentialist" assumptions mentioned in Section 1.1 imply the existence of a revealed preference ordering. And this ordering must satisfy Samuelson's (1952) independence axiom.

The set of preference orderings that satisfy the independence axiom on the space of relevant non-Archimedean random consequences is actually very large. Many such orderings, however, pay no attention to the interpretation of ϵ as an infinitesimal. In addition, recall that the only motivation which has been offered for non-Archimedean probabilities is to resolve the zero probability problem mentioned in Section 1.2. Now, were zero probabilities to be allowed, the problem they create is that the usual decision criteria generate behaviour sets that are too large. Accordingly, Section 4 proposes an additional and rather natural refinement axiom. This requires that in any decision tree whose chance moves have non-Archimedean probabilities differing only infinitesimally from those in an ordinary decision tree with real probabilities, behaviour with non-Archimedean probabilities should refine that with ordinary probabilities. It is then shown how this refinement axiom implies that the strict preference relation over non-Archimedean random consequences must refine the corresponding relation over ordinary random consequences. And, in the special case of RPFs, there must be a unique extension which can be regarded as a lexicographic hierarchy of preference orderings over ordinary random consequences.

Another part of the theory where the non-Archimedean field structure makes some difference is in the continuity condition set out in Section 5 as a necessary and sufficient condition for expected utility maximization. Within a general non-Archimedean field the construction used by Herstein and Milnor (1953) and others to determine the real values of a von Neumann–Morgenstern utility function is invalid, since it relies upon a continuity or "Archimedean" axiom. Nevertheless, the theory presented here is meant to extend the earlier standard theory of expected utility and subjective probability, so it seems reasonable to retain this Archimedean axiom for the subspace of decision trees with real-valued positive probabilities. In combination with the earlier axioms concerning behaviour in decision trees, this *restricted continuity* axiom implies that, when probabilities are described by RPFs, all the preference orderings making up the lexicographic hierarchy described in Section 4 can be represented by the expected value of the same von Neumann–Morgenstern utility function. A particular feature of the new lexicographic expected utility criterion is that it generalizes the standard criterion presented by Blume, Brandenburger and Dekel (1991a, b). For example, it allows and even requires expected utility hierarchies of differing lengths to be compared.

Now, the expectation of any real-valued utility function with respect to some non-Archimedean probability distribution is itself a member of the non-Archimedean ordered field introduced in Section 2. As Section 6 shows, the new lexicographic expected utility criterion is equivalent to maximizing non-Archimedean expected utility with respect to the ordering of the field $\Re(\epsilon)$ in which RPFs take their values.

Finally, Section 7 gathers together all the consequentialist axioms and their implications that have been set out in previous sections. They imply the existence of a unique class of cardinally equivalent von Neumann–Morgenstern utility functions such that consequentialist behaviour must be "non-Archimedean" Bayesian rational in the sense that it maximizes the lexicographic expected utility preference criterion set out in Section 4. Moreover, any behaviour which is non-Archimedean Bayesian rational in this sense will satisfy the consequentialist axioms. This is the main theorem of the paper.

2. Non-Archimedean Probabilities

2.1. General Non-Archimedean Ordered Fields

An *ordered field* $\langle \mathbb{F}, +, \cdot, 0, 1, > \rangle$ is a set $\mathbb{F}$ together with: (i) the two algebraic operations $+$ (addition) and $\cdot$ (multiplication); (ii) the two corresponding identity elements 0 and 1; (iii) the binary relation $>$ which is a total order of $\mathbb{F}$ satisfying $1 > 0$. The set $\mathbb{F}$ must be closed under its two algebraic operations. The usual properties of real number arithmetic also 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 $x \in \mathbb{F}$ must have both an additive inverse $-x$ and a multiplicative inverse $1/x$, except that $1/0$ is undefined. The order must be such that $y > z \iff y - z > 0$, while the set of positive elements in $\mathbb{F}$ must be closed under both addition and multiplication. Both the real line and the rationals are important and obvious examples of ordered fields. The set $\mathcal{Q}(\sqrt{2})$ of real numbers expressible in the form $a + b\sqrt{2}$ for some pair of rationals $a, b \in \mathcal{Q}$ is a somewhat less familiar example.

Any ordered field $\mathbb{F}$ has positive integer elements $n = 1, 2, \ldots$ which can be found by forming the sums $n = 1 + 1 + \ldots + 1$ of n copies of the element $1 \in \mathbb{F}$. Then $\mathbb{F}$ is said to be "Archimedean" if, given any $x > 0$ in $\mathbb{F}$, there exists such a positive integer n for which $n\,x > 1$. So any non-Archimedean ordered field must have at least one *positive infinitesimal* ϵ with the property that $n\,\epsilon \leq 1$ for all positive integers n.

The introduction claimed that the set of real-valued probabilities needs extending for a fully satisfactory decision theory. In future, therefore, $\mathbb{F}$ will always be a non-Archimedean ordered field that extends the real line $\Re$. Now, for any positive $r \in \Re$, there exists an ordinary positive integer n satisfying $1/n < r \in \mathbb{F}$. Any positive infinitesimal $\epsilon \in \mathbb{F}$ must therefore satisfy $\epsilon \leq 1/n < r$, and so must be smaller than any positive real number.

For any $x \in \mathbb{F}$, let $|x|$ denote x if $x \geq 0$ and $-x$ if $x < 0$. Say that any $x \in \mathbb{F}$ is *infinitesimal* if $|x| < r$ for all (small) real $r > 0$, and that x is *finite* if $|x| < r$ for some (large enough) real r. Any $x \in \mathbb{F}$ is said to be *infinite* if and only if it is not finite. Any non-zero $x \in \mathbb{F}$ is therefore infinitesimal if and only if $1/x$ is infinite.

Next, given any finite $x \in \mathbb{F}$, the two sets $\{ r \in \Re \mid r > x \}$ and $\{ r \in \Re \mid r \leq x \}$ partition the real line $\Re$, and so form a Dedekind cut. The usual properties of the real line then ensure that there is a unique ${}^0x \in \Re$, called the *real part* of x, defined by

$$ {}^0x := \inf \{ r \in \Re \mid r > x \} = \sup \{ r \in \Re \mid r \leq x \}. $$

For any real $r > 0$, note that $x - r < {}^0x < x + r$. Define ${}^\epsilon x := x - {}^0x$. Then $|{}^\epsilon x| < r$ for all real $r > 0$, so ${}^\epsilon x$ is infinitesimal. Since $x = {}^0x + {}^\epsilon x$, it is natural to call ${}^\epsilon x$ the *infinitesimal part* of x.

2.2. An Elementary Non-Archimedean Ordered Field

Since decision and game theory would seem to require some non-Archimedean ordered field $\mathbb{F}$ containing $\Re$ as a subfield, it is natural to explore the simplest such field. This must contain at least one positive infinitesimal ϵ. So no candidate for the field $\mathbb{F}$ can possibly be simpler than the one that results from appending ϵ to $\Re$, and then closing the resulting set $\Re \cup \{\epsilon\}$ under the operations of addition, subtraction, multiplication, and division except by zero. The result of this closure is a field denoted by $\Re(\epsilon)$ that has been discussed by Robinson (1973, p. 88–9) in particular. Its members are all the "rational" functions which can be expressed as ratios

$$ f(\epsilon) = \frac{A(\epsilon)}{B(\epsilon)} = \frac{a_0 + a_1\,\epsilon + a_2\,\epsilon^2 + \cdots + a_n\,\epsilon^n}{b_0 + b_1\,\epsilon + b_2\,\epsilon^2 + \cdots + b_m\,\epsilon^m} = \frac{\sum_{i=0}^n a_i\,\epsilon^i}{\sum_{i=0}^m b_i\,\epsilon^i} \tag{1} $$

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.

Now, one can simplify (1) by successively: (i) eliminating any leading zeros $a_0 = a_1 = \ldots = a_{k-1} = b_0 = b_1 = \ldots = b_{j-1} = 0$; (ii) dividing both numerator and denominator by the leading non-zero coefficient b_j of the denominator; (iii) cancelling any positive powers of ϵ that are common to all terms of both numerator and denominator and relabelling the coefficients a_i, b_i accordingly. The result is that any $f(\epsilon) \in \Re(\epsilon)$ given by (1) gets put into its *normalized form*

$$ f(\epsilon) = \frac{\sum_{i=k}^n a_i\,\epsilon^i}{\epsilon^j + \sum_{i=j+1}^m b_i\,\epsilon^i} \tag{2} $$

for some integers $j, k, m, n \geq 0$ such that $j = 0$ or $k = 0$ (or both). Moreover $a_k \neq 0$ unless $f(\epsilon) = 0$. Note too that each real number $r \in \Re$ can be simply expressed in the form (2) by writing $r = r/1 \in \Re(\epsilon)$, with $j, k, m, n = 0$ and $a_0 = r$.

It remains to be shown that $\Re(\epsilon)$ really is a non-Archimedean ordered field. The binary relation $>$ will be defined so that $y > z \iff y - z > 0$ where, for any $x = f(\epsilon) \in \Re(\epsilon)$ in its normalized form (2), one has $f(\epsilon) > 0$ if and only if $a_k > 0$. Recalling that ϵ is intended to be an infinitesimal, this condition is entirely natural because it is equivalent to having $f(r)$, the corresponding real-valued rational function of the real variable r, be positive for all small positive r. For later reference, note that $>$ is effectively a lexicographic relation. For if $f(\epsilon)$ is given by (1), then $f(\epsilon) > 0$ if and only if $a_k/b_j > 0$, where a_k and b_j are the first non-zero coefficients of the numerator and denominator respectively.

From the above definition, it is easy to check that $>$ is asymmetric and transitive, while either $f(\epsilon) > 0$ or $f(\epsilon) < 0$ unless $f(\epsilon) = 0$. So $>$ is indeed a total order. And it is easy and routine to check that the corresponding set of positive elements is closed under addition and multiplication. Finally, $\Re(\epsilon)$ is non-Archimedean because $1 > 0$ and so the order $>$ defined above satisfies $1 - n\,\epsilon > 0$ for every positive integer n.

Let $f(0) \in \Re$ denote the value at $r = 0$ of the corresponding polynomial function $f(r)$ of the real variable r. Then it is easy to see that any $f(\epsilon) \in \Re(\epsilon)$ given by (2) is infinitesimal if and only if $k > j = 0$, so that $f(0) = 0$. And any such $f(\epsilon)$ is infinite if and only if $j > k = 0$, so that $f(0)$ is undefined. Finally, whenever $f(\epsilon) \in \Re(\epsilon)$ is finite, $f(0)$ is equal to its real part ${}^0 f(\epsilon)$.

Choosing a more appropriate algebraic field in which probabilities can be defined is an important first step. And for the purposes of this paper and some later work in decision and game theory, the field $\Re(\epsilon)$ appears to be rich enough. However, some further extensions of $\Re(\epsilon)$ are eventually going to be necessary so as to accommodate countably additive non-Archimedean probability measures, continuous strategy spaces, etc. — for example, the set $\Re^\infty(\epsilon)$ mentioned in Hammond (1994, p. 48), whose members are ratios of power series $\sum_{k=0}^{\infty} a_k\,\epsilon^k$ with real coefficients $a_0, a_1, a_2, \ldots$. Accordingly, much of the paper actually works with a general non-Archimedean ordered field $\mathbb{F}$ that extends $\Re(\epsilon)$. This field could be as rich as the whole non-standard real line ${}^*\Re$ of "hyperreals," or it could be one of many possible sub-fields of ${}^*\Re$, including $\Re^\infty(\epsilon)$.

2.3. A Minimal Positive Cone

Probabilities are always non-negative. Avoiding the zero probability problem requires them all to be positive. In fact, when requiring a probability to have some positive non-Archimedean value $p \in \mathbb{F}$, a more restrictive condition than $p > 0$ will be used. Probabilities will actually be given values in some convex *non-Archimedean positive cone* $\mathbb{P}$ — that is, in a set $\mathbb{P} \subset \mathbb{F}_+ := \{ x \in \mathbb{F} \mid x > 0 \}$ containing all the positive reals and at least one infinitesimal, such that $\mathbb{P}$ is closed under addition, multiplication and division. Of course, such a cone cannot be closed under subtraction. So it is natural to have $\mathbb{P}$ as a minimal set in $\mathbb{F}$ with these properties. After all, this whole line of research is about having a space of probabilities that is no larger than absolutely necessary. The members of such a minimal $\mathbb{P}$ will be described as *strongly positive*.

When $\mathbb{F} = \Re(\epsilon)$, let $\mathcal{P}(\epsilon)$ denote the set of all $f(\epsilon) \in \Re(\epsilon)$ given by (1) whose real coefficients a_i ($i = 0$ to I) and b_j ($j = 0$ to J) are all non-negative, while $f(\epsilon) \neq 0$. Equivalently, the numerator and denominator of $f(\epsilon)$ must each contain at least one positive real coefficient, and no negative coefficient. Obviously $\mathcal{P}(\epsilon)$ is the smallest set containing both the positive part of the real line and ϵ which is also closed under addition, multiplication and division. The set $\mathcal{P}(\epsilon)$ is therefore a convex non-Archimedean positive cone, implying that one can take $\mathbb{P} = \mathcal{P}(\epsilon)$.

2.4. Non-Archimedean Probabilities

Let Ω be a non-empty sample space. Note carefully that Ω is not required to be finite. Let F any non-empty finite subset of Ω. A $\mathbb{P}$-*probability on* (or *with support*) F is a mapping $p(\cdot) : 2^\Omega \to \mathbb{P} \cup \{0\}$ which is defined on the domain 2^Ω of all subsets of Ω, and satisfies the three axioms:

(i) $p(E) \in \mathbb{P}$ whenever $E \cap F \neq \emptyset$;

(ii) $p(E) = 1$ whenever $F \subset E \subset \Omega$;

(iii) $p(E \cup E') = p(E) + p(E')$ whenever $E, E' \subset \Omega$ are such that $E \cap E' = \emptyset$.

From axioms (ii) and (iii) it follows that, whenever $E \cap F = \emptyset$, then

$$1 = p(E \cup F) = p(E) + p(F) = p(E) + 1$$

and so $p(E) = 0$. Actually, apart from (i), these are the usual axioms of probability theory, but with $p(E)$ taking values in $\mathbb{P} \cup \{0\}$ instead of $\Re_+$. However, axiom (i) strengthens the usual condition that $p(E) \geq 0$ for all $E \subset F$. To justify this strengthening, recall that the only reason for introducing extended and non-Archimedean probabilities has been the need to overcome the zero probability problem in decision and game theory. Note too that $p(E)$ is allowed to be an arbitrarily small strongly positive infinitesimal in $\mathbb{P}$.

Let $\Delta^0(F; \mathbb{P})$ denote the set of all such $\mathbb{P}$-probabilities with support F. This set is obviously an extension of the set $\Delta^0(F)$ of ordinary probability distributions with support F. Let $\Delta(\Omega; \mathbb{P})$ denote the set of all $\mathbb{P}$-probabilities that belong to $\Delta^0(F; \mathbb{P})$, for some finite $F \subset \Omega$. Thus all the members of $\Delta^0(F; \mathbb{P})$ definitely have F as their support, whereas each member of $\Delta(\Omega; \mathbb{P})$ has a support which can be any finite subset of Ω.

As usual with finitely supported probability distributions, any $p(\cdot) \in \Delta^0(F; \mathbb{P})$ is completely determined by its values $p(\{\omega\})$ on the singleton subsets $\{\omega\}$ ($\omega \in F$). With the customary slight abuse of notation, write these values as $p(\omega)$ ($\omega \in F$), all of which must be strongly positive. Moreover, it must be true that $p(\omega) = {}^0p(\omega) + \eta(\omega)$, where the real part ${}^0p(\cdot)$ is an ordinary probability distribution in $\Delta(F)$, while each $\eta(\omega)$ is infinitesimal in $\mathbb{F}$, and also $\sum_{\omega \in F} \eta(\omega) = 0$. For any $\omega \in F$ it is necessary to have ${}^0p(\omega) > 0$ or $\eta(\omega) \in \mathbb{P}$ (or both).

In the special case when $\mathbb{P} = \mathcal{P}(\epsilon)$, such non-Archimedean probabilities will be called *rational probability functions* (or RPFs). Let $\Delta^0(F; \epsilon)$ denote the set of all such RPFs with support F. And let $\Delta(\Omega; \epsilon)$ denote the set of all RPFs that belong to $\Delta^0(F; \epsilon)$, for some finite $F \subset \Omega$.

Members of $\Delta^0(F; \epsilon)$ have probabilities given, for all $\omega \in F$, by rational functions

$$p(\omega; \epsilon) = \frac{A(\omega; \epsilon)}{B(\omega; \epsilon)} = \frac{a_0(\omega) + a_1(\omega)\,\epsilon + a_2(\omega)\,\epsilon^2 + \cdots + a_{I(\omega)}(\omega)\epsilon^{I(\omega)}}{b_0(\omega) + b_1(\omega)\,\epsilon + b_2(\omega)\,\epsilon^2 + \cdots + b_{J(\omega)}(\omega)\epsilon^{J(\omega)}} = \frac{\sum_{i=0}^{I(\omega)} a_i(\omega)\,\epsilon^i}{\sum_{j=0}^{J(\omega)} b_j(\omega)\,\epsilon^j}$$

of the form (1). Here the coefficients $a_i(\omega)$ ($i = 0$ to $I(\omega)$) and $b_j(\omega)$ ($j = 0$ to $J(\omega)$) are all non-negative real numbers and, for any $\omega \in F$, neither all the $a_i(\omega)$ nor all the $b_j(\omega)$ are zero. Now, the finite collection $B(\omega; \epsilon)$ ($\omega \in F$) of polynomials has a positive lowest common denominator which will be written as $L(\epsilon) := \sum_{h=0}^{H} \ell_h\,\epsilon^h$, where each $\ell_h \geq 0$. For all $\omega \in F$ one has

$$p(\omega;\epsilon) = \frac{L(\omega;\epsilon)}{L(\epsilon)} = \frac{\sum_{h=0}^{H} \ell_h(\omega)\,\epsilon^h}{\sum_{h=0}^{H} \ell_h\,\epsilon^h}, \text{ where } L(\omega;\epsilon) := \frac{A(\omega;\epsilon)\,L(\epsilon)}{B(\omega;\epsilon)}. \tag{3}$$

Moreover, $\ell_h(\omega) \geq 0$ for all $\omega \in \Omega$ and for $h = 0, 1, 2, \ldots, H$. After repeating the operations used to obtain the normalized form (2) of a general rational function (1) — i.e., eliminating any leading zeros, cancelling redundant powers of ϵ, and dividing by the leading coefficient of the denominator — one obtains the *normalized form*

$$p(\omega;\epsilon) = \frac{p_0(\omega) + \sum_{h=1}^{H} \ell_h(\omega)\,\epsilon^h}{1 + \sum_{h=1}^{H} \ell_h\,\epsilon^h}. \tag{4}$$

Here $\sum_{\omega \in F} p_0(\omega) = 1$ and $\sum_{\omega \in F} \ell_h(\omega) = \ell_h$ for $h = 1$ to H. Moreover $p_0(\omega)$ must be equal to the real part ${}^0p(\omega;\epsilon)$ of $p(\omega;\epsilon)$, for all $\omega \in F$. In particular, $p_0(\cdot)$ is an ordinary probability distribution in $\Delta(F)$.

An alternative form of (4) will be used in Section 4 below. This comes from dropping any terms for which $\ell_h = 0$, while letting K denote the remaining set of those integers h with $\ell_h > 0$. Then (4) becomes

$$p(\omega;\epsilon) = \frac{\sum_{k \in K} \ell_k\,p_k(\omega)\,\epsilon^k}{\sum_{k \in K} \ell_k\,\epsilon^k} \tag{5}$$

where $p_k(\omega) := \ell_k(\omega)/\ell_k$ for $k \in K$ and all $\omega \in F$. Then $0 \in K$, $\ell_0 = 1$, and $p_k(\cdot) \in \Delta(F)$ for all $k \in K$.

Finally, note that (5) can be regarded as specifying probabilities

$$p^*(k,\omega;\epsilon) := \frac{\ell_k\,p_k(\omega)\,\epsilon^k}{\sum_{j \in K} \ell_j\,\epsilon^j}$$

as functions of ϵ on the extended sample space $\mathcal{N} \times \Omega$, where $\mathcal{N}$ denotes the set of non-negative integers, to be interpreted as various orders in a lexicographic hierarchy (cf. Blume *et al.*, 1991a, b). This is equivalent to compounding the ordinary conditional distributions $p^*(\omega|k) := p_k(\omega)$ (all $\omega \in F$ and $k \in K$) with the particular RPF $\ell(\cdot;\epsilon) \in \Delta(\mathcal{N};\epsilon)$ given by

$$\ell(k;\epsilon) := p^*(\{k\} \times \Omega;\epsilon) = \frac{\ell_k\,\epsilon^k}{\sum_{j \in K} \ell_j\,\epsilon^j} \quad (\text{all } k \in K).$$

2.5. *Independence*

Suppose that the finite non-empty sample space Ω is the n-fold Cartesian product $\prod_{s=1}^n \Omega^s$. Suppose that, for $s = 1$ to n, there are $\mathbb{P}$-probability distributions $p^s(\cdot) \in \Delta^0(F^s; \mathbb{P})$ with respective finite supports $F^s \subset \Omega^s$. Then these n distributions are said to be *independent* if their joint distribution has support $F = \prod_{s=1}^n F^s$ and is described by the unique $\mathbb{P}$-probability distribution $p(\cdot) \in \Delta(F; \mathbb{P})$ that satisfies

$$p(\prod_{s=1}^n E^s) = \prod_{s=1}^n p^s(E^s) \quad \text{whenever } E^s \subset F^s \ (s = 1, 2, \ldots, n).$$

This corresponds to the strongest of the three definitions of independence in Blume *et al.* (1991a) and in Hammond (1994). And it is the only one of the three allowing the joint distribution to be inferred uniquely from the independent marginal distributions.

2.6. *Non-Archimedean Conditional Probabilities*

Suppose that $p(\cdot) \in \Delta^0(F; \mathbb{P})$ is any $\mathbb{P}$-probability distribution on the finite subset $F \subset \Omega$. Then the non-Archimedean conditional probability

$$P(E|E') := p(E)/p(E') \in \mathbb{P} \tag{6}$$

is certainly well defined whenever $\emptyset \neq E \subset E' \subset F$. The zero probability problem has therefore been resolved.

Consider the case when $\mathbb{P} = \mathcal{P}(\epsilon)$ and $p(\cdot; \epsilon) \in \Delta^0(F; \epsilon)$ is any RPF on the finite subset $F \subset \Omega$. Then the non-Archimedean conditional probability (6) takes the form

$$P(E|E'; \epsilon) := p(E; \epsilon)/p(E'; \epsilon) \in \mathcal{P}(\epsilon) \tag{7}$$

where $p(E; \epsilon)$ and $p(E'; \epsilon)$ are rational functions in $\mathcal{P}(\epsilon)$, and so therefore is $P(E|E'; \epsilon)$. Since ϵ is an infinitesimal, it is tempting in this case to consider the ordinary conditional probabilities that are defined by the limit

$$P(E|E') := \lim_{r \to 0+} P(E|E'; r), \tag{8}$$

of the positive real-valued rational function $P(E|E'; r)$ as r tends to zero through ordinary positive real values. The conditional probabilities $P(E|E')$ ($\emptyset \neq E \subset E' \subset F$) then form a *complete conditional probability system* (or CCPS) on F, of the kind studied by Rényi (1955, 1970), Lindley (1965), and Myerson (1986), amongst others. Each such CCPS is obviously represented by an equivalence class of RPFs, where any two RPFs $p(\cdot;\epsilon), q(\cdot;\epsilon) \in \Delta^0(F;\epsilon)$ are to be regarded as equivalent if and only if

$$\lim_{r \to 0+} p(E;r)/p(E';r) = \lim_{r \to 0+} q(E;r)/q(E';r) \text{ whenever } \emptyset \neq E \subset E' \subset F. \tag{9}$$

Moreover, as pointed out in Hammond (1994), each such equivalence class of RPFs can be represented by a single *conditional rational probability function* satisfying the property that

$$p(\omega;\epsilon) = \frac{p_{k(\omega)}(\omega)\,\epsilon^{k(\omega)}}{\sum_{k=0}^{K} \epsilon^k} \tag{10}$$

for all $\omega \in F$, where $p_k \in \Delta^0(F_k)$ are ordinary probability distributions whose supports

$$F_k = \{\, \omega \in F \mid p_k(\omega) > 0 \,\} = \{\, \omega \in F \mid k(\omega) = k \,\} \quad (k = 0 \text{ to } K)$$

form a partition of F. However, considering only such CCPSs loses a lot of valuable relevant information. For, as discussed in Section 1.2, the joint distribution of several independent random variables is only uniquely determined from the marginal distributions when non-Archimedean probabilities are represented by RPFs in general form, with values in an algebraic field.

3. Non-Archimedean Consequentialist Behaviour

3.1. Non-Archimedean Consequentialist Decision Trees

Let Y be a given domain of consequences which could conceivably occur. As remarked in Section 1.1, consequentialism in single-person decision theory considers dynamically consistent behaviour β in an almost unrestricted domain $\mathcal{T}(Y)$ of consequentialist finite decision trees. These are decision trees in the sense of Raiffa (1968), except that payoffs are not specified. Instead, any terminal node x in any tree $T \in \mathcal{T}(Y)$ has an associated consequence $y = \gamma(x)$ within the consequence domain Y.

This paper will consider instead the set $\mathcal{T}(Y;\mathbb{P})$ of $\mathbb{P}$-*consequentialist* finite decision trees. Each member of $\mathcal{T}(Y;\mathbb{P})$ is a collection

$$T = \langle N, N^*, N^0, X, N_{+1}(\cdot), n_0, \pi(\cdot|\cdot), \gamma(\cdot) \rangle \tag{11}$$

whose eight component parts are described and interpreted as follows:

(i) N is a non-empty finite set of *nodes* of the tree T, which is partitioned into the three disjoint sets N^*, N^0, and X having properties (ii)–(vii) below;

(ii) N^* is the (possibly empty) set of *decision nodes*;

(iii) N^0 is the (possibly empty) set of *chance nodes*;

(iv) X is the non-empty set of *terminal nodes*;

(v) $N_{+1} : N \twoheadrightarrow N$ is the *immediate successor* correspondence which, so that T really is indeed a tree, must satisfy:

(a) $\forall n \in N : \; n \notin N_{+1}(n)$;

(b) $\forall n \in N : \; N_{+1}(n) = \emptyset \iff n \in X$;

(c) $\forall n, n' \in N : \; N_{+1}(n) \cap N_{+1}(n') \neq \emptyset \iff n = n'$;

and must also generate an acyclic binary *successor* relation $\succ$ on N, defined by the property that $n' \succ n$ if and only if there exists a chain $n_1, n_2, \ldots, n_k$ in N such that $n_1 = n$, $n_k = n'$ and $n_{j+1} \in N_{+1}(n_j)$ for $j = 1$ to $k - 1$;

(vi) n_0 is the unique *initial node* in N satisfying $\forall n \in N : \; n_0 \notin N_{+1}(n)$;

(vii) at each chance node $n \in N^0$, there is a non-Archimedean conditional probability distribution $\pi(\cdot|n) \in \Delta^0(N_{+1}(n); \mathbb{P})$ over the set $N_{+1}(n)$ (whose members are to be interpreted as the possible chance moves from n to each succeeding node $n' \in N_{+1}(n)$);

(viii) $\gamma : X \to Y$ is the *consequence mapping*, indicating the consequence $\gamma(x)$ of reaching each terminal node x.

Most of this definition is identical to that in Hammond (1988a, b, Section 2). Natural nodes and sets of possible states of the world are not considered, since consideration of subjective probabilities has been left for later work. Otherwise, only the concept of $\mathbb{P}$-probability in part (vii) and the consequence mapping in part (viii) are different. The latter has been changed so that each terminal node x gives rise to just one sure consequence $\gamma(x) \in Y$, rather than to a probability distribution over Y. The change is just to make clear how all uncertainty is eventually resolved within a decision tree. The proof of the Refinement Lemma in Section 4.1 below contains an example illustrating this definition.

Where it is desirable to emphasize dependence on T, I shall write $N(T)$, $N^*(T)$, etc.

3.2. *Dynamically Consistent Behaviour*

As in the earlier work described in Section 1.1, it will now be assumed that *behaviour* is formally described by a correspondence β with a domain consisting of all pairs (T, n) satisfying $T \in \mathcal{T}(Y; \mathbb{P})$ and $n \in N^*(T)$. Its value is a *behaviour set* $\beta(T, n)$, which is a non-empty subset of the appropriate set $N_{+1}(T, n)$ of all decisions that are feasible at the decision node n of the tree T in the domain $\mathcal{T}(Y; \mathbb{P})$ of non-Archimedean consequentialist decision trees.

Now, given any decision tree $T \in \mathcal{T}(Y; \mathbb{P})$ and any fixed node $\bar{n}$ of T, there is a "subtree" of T or a *continuation from* $\bar{n}$ with

$$T(\bar{n}) := \bar{T} = \langle \bar{N}, \bar{N}^*, \bar{N}^0, \bar{X}, \bar{N}_{+1}(\cdot), \bar{n}_0, \bar{\pi}(\cdot|\cdot), \bar{\gamma}(\cdot) \rangle. \tag{12}$$

To define $T(\bar{n})$ explicitly, first let $N(n) := \{ n' \in N \mid n' \succ n \text{ or } n' = n \}$ denote the set of nodes in N which either succeed or coincide with n. Then $T(\bar{n})$ is the decision tree with the initial node $\bar{n}_0 := \bar{n}$, the set of nodes $\bar{N} := N(\bar{n})$ and with all of the other sets, correspondences, and functions $\bar{N}^*, \bar{N}^0, \bar{X}, \bar{N}_{+1}(\cdot), \bar{\pi}(\cdot|\cdot), \bar{\gamma}(\cdot)$ of (12) given by appropriate restrictions of $N^*, N^0, X, N_{+1}(\cdot), \pi(\cdot|\cdot), \gamma(\cdot)$ to the smaller set of nodes $\bar{N}$. From this definition it is obvious that $T(\bar{n})$ meets all the criteria set out in Section 3.1 above, and so is itself an non-Archimedean consequentialist decision tree in $\mathcal{T}(Y; \mathbb{P})$.

Suppose that n is any decision node of a tree $T \in \mathcal{T}(Y; \mathbb{P})$ with the set of nodes N. When the agent comes to make a decision at node n, the remaining decision problem is really sufficiently described by the tree $T(n)$. Behaviour at node n must therefore be described by $\beta(T(n), n)$ as well as by $\beta(T, n)$. Thus it will also be assumed that behaviour is *dynamically consistent* in the sense that $\beta(T, n) = \beta(T(n), n)$ at all decision nodes n of every tree $T \in \mathcal{T}(Y; \mathbb{P})$. This consistency condition is entirely natural since, as discussed in Hammond (1988a, b), it is satisfied even by a naïve agent who neglects changing tastes altogether. Of course, a naïve agent's actual behaviour usually departs from planned behaviour. Also, except in very special cases, naïve behaviour violates consequentialism, as defined below (cf. Hammond, 1976).

From now on only consistent behaviour will be considered, and will be called simply *behaviour*.

3.3. Non-Archimedean Consequences of Behaviour

Finally, it will be presumed that the domain of consequences has been specified broadly enough to capture everything that is relevant for the agent's behaviour. Then "consequentialism" means that in all decision trees in the domain, including the "continuation subtrees" of any tree, the agent's behaviour gives rise to a non-empty set of possible "chosen" random consequences which depends only upon the feasible set of random consequences. Before this can be formalized it is obviously necessary to see first what the feasible set and choice set are in any tree $T \in \mathcal{T}(Y; \mathbb{P})$.

Let $\Delta(Y; \mathbb{P})$ denote the set of all $\mathbb{P}$-probabilities when the sample space is taken to be the consequence domain Y. It turns out that, in any tree $T \in \mathcal{T}(Y; \mathbb{P})$, any kind of feasible behaviour must yield a random consequence belonging to some non-empty *feasible set* $F(T) \subset \Delta(Y; \mathbb{P})$. Moreover, behaviour β must yield a random consequence belonging to some non-empty *(revealed) choice set* $\Phi_\beta(T) \subset F(T)$. This is best demonstrated by using backward recursion to construct, in successively larger and larger subtrees $T(n)$ with earlier and earlier initial nodes n in $N(T)$, the corresponding pairs of subsets $F(T, n)$ and $\Phi_\beta(T, n)$ of $\Delta(Y; \mathbb{P})$. Moreover, backward induction can be used to show that

$$\emptyset \neq \Phi_\beta(T, n) \subset F(T, n) \subset \Delta(Y; \mathbb{P})$$

at every node $n \in N(T)$. Finally, of course, one works back to the initial node n_0 at which $F(T) = F(T, n_0)$ and $\Phi_\beta(T) = \Phi_\beta(T, n_0)$, so that

$$\emptyset \neq \Phi_\beta(T) \subset F(T) \subset \Delta(Y; \mathbb{P}).$$

The backward recursion and backward induction start at the terminal nodes $x \in X$ where only a single consequence $\gamma(x)$ is possible. Then

$$\emptyset \neq F(T, x) = \Phi_\beta(T, x) = \{\delta_{\gamma(x)}(\cdot)\} \subset \Delta(Y; \mathbb{P}) \tag{13}$$

where $\delta_{\gamma(x)}(\cdot)$ denotes the unique degenerate $\mathbb{P}$-probability distribution in $\Delta^0(\{\gamma(x)\}; \mathbb{P})$ with $\delta_{\gamma(x)}(\{\gamma(x)\}) = 1$. At previous nodes $n \in N \setminus X$, both $F(T, n)$ and $\Phi_\beta(T, n)$ are calculated from the values of $F(T, n')$ and $\Phi_\beta(T, n')$ at all nodes $n' \in N_{+1}(n)$ which immediately succeed n. There are two different cases — cf. Hammond (1988b, Section 4).

Case 1. At any decision node $n \in N^*$ one has

$$\emptyset \neq \Phi_\beta(T,n) := \bigcup_{n' \in \beta(T,n)} \Phi_\beta(T,n') \subset F(T,n) := \bigcup_{n' \in N_{+1}(n)} F(T,n') \subset \Delta(Y;\mathbb{P}). \tag{14}$$

Case 2. At any chance node $n \in N^0$ one has

$$\begin{aligned} \emptyset \neq \Phi_\beta(T,n) &:= \sum\nolimits_{n' \in N_{+1}(n)} \pi(n'|n)\, \Phi_\beta(T,n') \\ \subset F(T,n) &:= \sum\nolimits_{n' \in N_{+1}(n)} \pi(n'|n)\, F(T,n') \subset \Delta(Y;\mathbb{P}). \end{aligned} \tag{15}$$

These constructions (14) and (15) evidently work just as well in the set $\Delta(Y;\mathbb{P})$ as they did before in $\Delta(Y)$. The proof of the Refinement Lemma in Section 4.1 provides a simple example of how they work in practice.

3.4. Non-Archimedean Consequentialist Behaviour

The *non-Archimedean consequentialist choice axiom* requires behaviour β to reveal a unique *non-Archimedean consequence choice function* C_β with the property that

$$\Phi_\beta(T) = C_\beta(F(T)) \tag{16}$$

in all the trees $T \in \mathcal{T}(Y;\mathbb{P})$. In particular, the structure of the decision tree must be irrelevant to consequentialist behaviour, as long as the feasible set $F(T)$ of possible distributions stays the same.

3.5. Ordinality and Independence

So the three consequentialist axioms require that the behaviour sets $\beta(T,n)$: (i) are non-empty subsets of $N_{+1}(n)$ defined at every decision node n of every tree $T \in \mathcal{T}(Y;\mathbb{P})$; (ii) satisfy the consistency condition $\beta(T(n),n') = \beta(T,n')$ at all decision nodes n' of any subtree $T(n)$ of a tree $T \in \mathcal{T}(Y;\mathbb{P})$; (iii) satisfy the consequentialist choice axiom (16).

The arguments of Hammond (1988b, Sections 5 and 6) can now be applied virtually without change, and so will not be repeated here. They show first that consequentialist behaviour β reveals, not only a consequence choice function C_β, but also a corresponding *consequence preference ordering* R — i.e., a complete transitive binary relation — on the set $\Delta(Y;\mathbb{P})$ of $\mathbb{P}$-probabilities. Thus, for every $T \in \mathcal{T}(Y;\mathbb{P})$, one has

$$\Phi_\beta(T) = \{\, \lambda \in F(T) \mid \mu \in F(T) \Longrightarrow \lambda \; R \; \mu \,\}.$$

Second, on the space $\Delta(Y)$ of ordinary probability distributions, which is of course a subset of $\Delta(Y;\mathbb{P})$, it was shown that the *independence condition* is another implication of consequentialism. For all ordinary probability distributions $\lambda, \mu, \nu \in \Delta(Y)$ and all real numbers α with $0 < \alpha < 1$, independence requires that

$$\lambda \; R \; \mu \iff [\alpha\,\lambda + (1-\alpha)\,\nu] \; R \; [\alpha\,\mu + (1-\alpha)\,\nu]. \tag{17}$$

Exactly the same arguments now imply the *non-Archimedean independence condition* on $\Delta(Y;\mathbb{P})$. For all $\lambda, \mu, \nu \in \Delta(Y;\mathbb{P})$ this requires that

$$\lambda \; R \; \mu \iff \frac{\alpha\,\lambda + \theta\,\nu}{\alpha + \theta} \; R \; \frac{\alpha\,\mu + \theta\,\nu}{\alpha + \theta} \tag{18}$$

whenever $\alpha, \theta \in \mathbb{P}$, so that $\alpha\,(\alpha+\theta)^{-1}$ and $\theta\,(\alpha+\theta)^{-1}$ can be regarded as $\mathbb{P}$-probabilities over a pair of disjoint states.

In Hammond (1988b, Section 8) it is also shown that any preference ordering R over the set $\Delta(Y)$ satisfying the independence condition (17) generates consequentialist and dynamically consistent behaviour on the "almost unrestricted domain" of all finite consequentialist decision trees having strictly positive ordinary probabilities at all chance nodes. It is important to consider only trees in this almost unrestricted domain because, if a completely unrestricted domain of finite consequentialist decision trees were allowed instead, then the implication would be that all random consequences in each space $\Delta(Y)$ must be indifferent. Here, a corresponding result is true, but on the less restricted domain $\mathcal{T}(Y;\mathbb{P})$ of finite consequentialist decision trees having strongly positive non-Archimedean probabilities at all chance nodes. For any preference ordering R over the set $\Delta(Y;\mathbb{P})$ satisfying the new independence condition (18) will generate consequentialist and dynamically consistent behaviour on this domain. Once again, the proof is exactly the same as before.

4. Lexicographic Refinements

4.1. A Refinement Axiom

There are many different preference orderings R on $\Delta(Y;\mathbb{P})$ satisfying independence (18). Some of these are implausible, however. For consider the case when $\mathbb{P} = \mathcal{P}(\epsilon)$. Let $v : Y \to \Re$ be any von Neumann–Morgenstern utility function defined on the set Y of all possible sure consequences. Then, for each positive real number r, there is a corresponding ordering R_r with the property that

$$p\ R_r\ q \iff \sum\nolimits_{y\in Y} [p(y;r) - q(y;r)]\, v(y) \geq 0 \tag{19}$$

whenever $p, q \in \Delta(Y;\epsilon)$, where $p(y;r)$ and $q(y;r)$ denote the real numbers obtained by regarding p and q as functions of the real variable r instead of the infinitesimal ϵ. The preferences defined by (19) obviously satisfy independence (18), as well as the restricted continuity condition to be considered later. Yet such orderings pay no attention to the intended interpretation of ϵ as an infinitesimal. A further assumption is needed.

As explained in Hammond (1994), the reason for introducing probabilities with values in a general non-Archimedean cone $\mathbb{P}$ is to refine consequentialist behaviour in trees having zero probabilities at some chance nodes, while leaving unaffected such behaviour in trees having positive probabilities at all chance nodes. Now, given any non-Archimedean tree $T \in \mathcal{T}(Y;\mathbb{P})$ as in Section 3.1, let ${}^0T \in \mathcal{T}(Y)$ be the corresponding ordinary tree obtained when every non-Archimedean probability $\pi(n'|n) \in \mathbb{P}$ ($n \in N^0$, $n' \in N_{+1}(n)$) in T is replaced by its corresponding real part ${}^0\pi(n'|n) \in \Re$. Also, to ensure that ${}^0\pi(n'|n) > 0$ throughout ${}^0T$, remove the entire subtree ${}^0T(n')$ from ${}^0T$ whenever ${}^0\pi(n'|n) = 0$ for some node $n' \in N_{+1}(n)$ immediately succeeding the chance node $n \in N^0$.

The *refinement axiom* then requires that at any decision node $n \in N^*({}^0T) \subset N^*(T)$, one must have $\beta(T,n) \subset \beta({}^0T,n)$. In particular, if $n \in N^*({}^0T)$ is such that $\beta({}^0T,n) = \{n'\}$ for some unique $n' \in N_{+1}(n)$, then it must be true that $\beta(T,n) = \{n'\}$ as well. Actually, it will be enough to assume this weaker condition in what follows. The main implication of this refinement axiom is the following Lemma, whose conclusion could almost have been stated as an alternative to the axiom.

Refinement Lemma. *In combination with the consequentialist axioms of Section 3.5, the refinement axiom implies that, whenever $p, q \in \Delta(Y;\mathbb{P})$ and ${}^0p\ P\ {}^0q$, then $p\ P\ q$.*

PROOF: First, let

$$Y_p := \{ y \in Y \mid p(y) > 0 \} \quad \text{and} \quad Y_q := \{ y \in Y \mid q(y) > 0 \}$$

denote the finite supports of the two distributions $p, q \in \Delta(Y; \mathbb{P})$. Then consider the particular finite decision tree $T \in \mathcal{T}(Y; \mathbb{P})$ with initial node n_0, which is the only decision node, while the other nodes satisfy:

$$N_{+1}(n_0) = N^0 = \{ n_p, n_q \} \quad \text{and} \quad X = X_p \cup X_q$$

$$\text{where } X_p := N_{+1}(n_p) = \{ x_p(y) \mid y \in Y_p \} \quad \text{and} \quad X_q := N_{+1}(n_q) = \{ x_q(y) \mid y \in Y_q \}.$$

Suppose too that the probabilities and consequences in T are given by

$$\pi(x_p(y)|n_p) = p(y) \quad \text{and} \quad \gamma(x_p(y)) = y \text{ (all } y \in Y_p);$$

$$\pi(x_q(y)|n_q) = q(y) \quad \text{and} \quad \gamma(x_q(y)) = y \text{ (all } y \in Y_q).$$

Then the backward recursion construction of Section 3.3 shows that

$$\Phi_\beta(T, x_p(y)) = F(T, x_p(y)) = \{\delta_y\} \text{ (all } y \in Y_p);$$

$$\text{and} \quad \Phi_\beta(T, x_q(y)) = F(T, x_q(y)) = \{\delta_y\} \text{ (all } y \in Y_q)$$

at the terminal nodes, while

$$\Phi_\beta(T, n_p) = F(T, n_p) = \sum\nolimits_{x \in X_p} \pi(x|n_p)\, F(T, x) = \sum\nolimits_{y \in Y_p} p(y)\, \{\delta_y\} = \{p\};$$

$$\Phi_\beta(T, n_q) = F(T, n_q) = \sum\nolimits_{x \in X_q} \pi(x|n_q)\, F(T, x) = \sum\nolimits_{y \in Y_q} q(y)\, \{\delta_y\} = \{q\}.$$

at the chance nodes. So finally the feasible set is

$$F(T) = F(T, n_0) = F(T, n_p) \cup F(T, n_q) = \{p\} \cup \{q\} = \{ p, q \} \subset \Delta(Y; \mathbb{P}).$$

Also the revealed choice set of consequences is $\Phi_\beta(T) = \Phi_\beta(T, n_0)$, where

$$p \in \Phi_\beta(T, n_0) \iff n_p \in \beta(T, n_0) \quad \text{and} \quad q \in \Phi_\beta(T, n_0) \iff n_q \in \beta(T, n_0).$$

Now ${}^0T$ differs from T only in replacing the probabilities $p(y)$ and $q(y)$ by their respective real parts ${}^0p(y)$, ${}^0q(y)$, while also excluding altogether any terminal nodes $x_p(y)$ or $x_q(y)$ for which ${}^0p(y) = 0$ or ${}^0q(y) = 0$, as appropriate. So a similar construction in ${}^0T$ shows that $F({}^0T) = \{ {}^0p, {}^0q \} \subset \Delta(Y)$. Then the hypothesis ${}^0p\ P\ {}^0q$ implies $\Phi_\beta({}^0T) = \{{}^0p\}$. This obviously requires $\beta({}^0T, n_0) = \{n_p\}$. But then the refinement axiom, even in its weak form, implies $\beta(T, n_0) = \{n_p\}$. Therefore $\Phi_\beta(T) = \{p\}$, implying that $p\ P\ q$. ∎

This lemma already gives the beginnings of a lexicographic preference criterion. It is incomplete, however, because in the case when ${}^0p\ I\ {}^0q$ nothing yet has been said about what higher criterion determines whether or not $p\ P\ q$.

4.2. Lexicographic Preferences

This section will consider RPFs with values in $\mathcal{P}(\epsilon)$ and show that there must be a complete lexicographic preference criterion in this special case. To this end, let

$$p(y;\epsilon) = \frac{\sum_{i\in I} P_i \, p_i(y)\, \epsilon^i}{\sum_{i\in I} P_i \, \epsilon^i} \quad \text{and} \quad q(y;\epsilon) = \frac{\sum_{j\in J} Q_j \, q_j(y)\, \epsilon^j}{\sum_{j\in J} Q_j \, \epsilon^j} \tag{20}$$

be any two RPFs $p, q \in \Delta(Y;\epsilon)$ that have been expressed in the form (5). In particular, it must be true that $0 \in I \cap J$ and $P_0 = Q_0 = 1$, while $P_i > 0$, $p_i(\cdot) \in \Delta(Y)$ (all $i \in I$) and $Q_j > 0$, $q_j(\cdot) \in \Delta(Y)$ (all $j \in J$). Now these two RPFs can be given the common denominator

$$\sum_{k\in K} R_k \, \epsilon^k := \sum_{i\in I} P_i \, \epsilon^i \times \sum_{j\in J} Q_j \, \epsilon^j$$

where $K := I + J$ (as the sum of the two sets of integers) and $R_k := \sum_{(i,j)\in H_k} P_i \, Q_j$ with $H_k := \{\, (i,j) \in I \times J \mid i + j = k \,\}$ (all $k \in K$). These definitions, together with the previously specified properties of I, J, and P_i $(i \in I)$, Q_j $(j \in J)$, also imply that $0 \in K$, while $R_0 = 1$ and $R_k > 0$ $(k \in K)$. Then one has

$$p(y;\epsilon) = \frac{\sum_{i\in I} P_i \, p_i(y)\, \epsilon^i \times \sum_{j\in J} Q_j \, \epsilon^j}{\sum_{k\in K} R_k \, \epsilon^k} = \frac{\sum_{k\in K} \sum_{(i,j)\in H_k} P_i \, p_i(y) \, Q_j \, \epsilon^k}{\sum_{k\in K} R_k \, \epsilon^k}. \tag{21}$$

Now define $p_k^*(\cdot), q_k^*(\cdot) \in \Delta(Y)$ (all $k \in K$) so that

$$p_k^*(y) := \sum_{(i,j)\in H_k} \frac{P_i \, Q_j}{R_k} \, p_i(y) \quad \text{and} \quad q_k^*(y) := \sum_{(i,j)\in H_k} \frac{P_i \, Q_j}{R_k} \, q_j(y). \tag{22}$$

With this definition, (21) implies that

$$p(y;\epsilon) = \frac{\sum_{k\in K} R_k \, p_k^*(y)\, \epsilon^k}{\sum_{k\in K} R_k \, \epsilon^k} \quad \text{and similarly} \quad q(y;\epsilon) = \frac{\sum_{k\in K} R_k \, q_k^*(y)\, \epsilon^k}{\sum_{k\in K} R_k \, \epsilon^k}.$$

So, after dropping the function arguments from the notation, one has

$$p = \frac{\sum_{k\in K} R_k \, p_k^* \, \epsilon^k}{\sum_{k\in K} R_k \, \epsilon^k} \quad \text{and} \quad q = \frac{\sum_{k\in K} R_k \, q_k^* \, \epsilon^k}{\sum_{k\in K} R_k \, \epsilon^k}. \tag{23}$$

Next, for any integer $m \in K$, define the following three subsets of K:

$$K_{<m} := \{\, k \in K \mid k < m \,\}; \quad K_{>m} := \{\, k \in K \mid k > m \,\}; \quad K_{\geq m} := \{\, k \in K \mid k \geq m \,\}.$$

Furthermore, define

$$R^m(\epsilon) := \sum\nolimits_{k \in K_{\geq m}} R_k\, \epsilon^k, \quad p^m := \frac{\sum_{k \in K_{\geq m}} R_k\, p_k^*\, \epsilon^k}{R^m(\epsilon)}, \quad q^m := \frac{\sum_{k \in K_{\geq m}} R_k\, q_k^*\, \epsilon^k}{R^m(\epsilon)}. \tag{24}$$

Then one can rewrite (23) as

$$p = \frac{\sum_{k \in K_{<m}} R_k\, p_k^*\, \epsilon^k + R^m(\epsilon)\, p^m}{\sum_{k \in K_{<m}} R_k\, \epsilon^k + R^m(\epsilon)}; \quad q = \frac{\sum_{k \in K_{<m}} R_k\, q_k^*\, \epsilon^k + R^m(\epsilon)\, q^m}{\sum_{k \in K_{<m}} R_k\, \epsilon^k + R^m(\epsilon)}.$$

In case $p_k^*\ I\ q_k^*$ for $k \in K_{<m}$, it follows from repeated application of the independence condition (18) that $p\ P\ q \iff p^m\ P\ q^m$.

Now, dividing by $R_m\, \epsilon^m$ all the terms of both numerator and denominator in the definitions of p^m, q^m in (24) gives

$$p^m = \frac{p_m^* + \sum_{k \in K_{>m}} R_k\, p_k^*\, \epsilon^{k-m}/R_m}{1 + \sum_{k \in K_{>m}} R_k\, \epsilon^{k-m}/R_m} \quad \text{and} \quad q^m = \frac{q_m^* + \sum_{k \in K_{>m}} R_k\, q_k^*\, \epsilon^{k-m}/R_m}{1 + \sum_{k \in K_{>m}} R_k\, \epsilon^{k-m}/R_m}.$$

The real parts of these probability distributions are therefore ${}^0p^m = p_m^*$ and ${}^0q^m = q_m^*$. Combining the above results with the refinement lemma then shows that, in case $p_k^*\ I\ q_k^*$ for all $k \in K_{<m}$, one has

$$p_m^*\ P\ q_m^* \iff {}^0p^m\ P\ {}^0q^m \Longrightarrow p^m\ P\ q^m \iff p\ P\ q$$

and so $p_m^*\ P\ q_m^* \Longrightarrow p\ P\ q$. Finally, in case $p_k^*\ I\ q_k^*$ for all $k \in K$, repeated application of the independence condition (18) shows that $p\ I\ q$. So it has been proved that

Theorem. *Suppose that the consequentialist axioms of Section 3.5 and the refinement axiom of Section 4.1 are all satisfied. Then, for all pairs of RPFs $p, q \in \Delta(Y; \epsilon)$ given by (20), one has*

$$p\ P\ q \iff \langle p_k^* \rangle_{k \in K}\ P_L\ \langle q_k^* \rangle_{k \in K}$$

where $\langle p_k^* \rangle_{k \in K}, \langle q_k^* \rangle_{k \in K} \in \Delta^K(Y)$ are the lexicographic hierarchies of probability distributions whose members are given by (22), and where P_L is the lexicographic preference criterion defined by

$$\langle p_k^* \rangle_{k \in K}\ P_L\ \langle q_k^* \rangle_{k \in K} \iff \exists m \in K : p_k^*\ I\ q_k^*\ (\text{all } k \in K_{<m})\ \&\ p_m^*\ P\ q_m^*.$$

5. Continuity and von Neumann–Morgenstern Utility

5.1. Continuous Preferences

In Section 3.5 it was claimed that consequentialist behaviour in trees $T \in \mathcal{T}(Y; \mathbb{P})$ must maximize a preference ordering R on $\Delta(Y; \mathbb{P})$ satisfying the non-Archimedean independence axiom (18). Because (18) implies the ordinary independence condition (17), two of Jensen's (1967) three axioms which imply expected utility maximization are satisfied. But discontinuous lexicographic preferences are still possible, even in trees with strictly positive ordinary probabilities at chance nodes. In single person decision theory, the motivation for avoiding such discontinuities is not entirely clear, beyond analytical convenience and a feeling that continuity is anyway rather natural. In the case of n-person game theory, however, continuity of behaviour as common expectations vary is crucial for the existence of equilibrium in general games.

To derive expected utility maximization, an axiom of continuity as probabilities vary — or what some have called an "Archimedean" axiom — will be added. The following was included as Jensen's (1967) third and last axiom, modifying Herstein and Milnor's (1953) simplification of the original von Neumann and Morgenstern (1944) formulation:

RESTRICTED CONTINUITY AXIOM. *If $\lambda, \mu, \nu \in \Delta(Y)$ with $\lambda\ P\ \mu$ and $\mu\ P\ \nu$, then there exist real numbers α and θ with $0 < \alpha < \theta < 1$ such that*

$$[(1-\alpha)\,\lambda + \alpha\,\nu]\ P\ \mu \quad \text{and} \quad \mu\ P\ [(1-\theta)\,\lambda + \theta\,\nu].$$

As Jensen (1967) shows, this assumption, together with the fact that R is a preference ordering on $\Delta(Y)$ satisfying the independence axiom (17), implies the existence of a real-valued von Neumann–Morgenstern utility function (NMUF) v on Y such that

$$\lambda\ R\ \mu \iff \mathbb{E}_\lambda\, v \geq \mathbb{E}_\mu\, v \tag{25}$$

for all $\lambda, \mu \in \Delta(Y)$, where

$$\mathbb{E}_\lambda\, v := \sum\nolimits_{y \in Y} \lambda(y)\, v(y) \tag{26}$$

denotes the expected value of v with respect to λ, and $\mathbb{E}_\mu\, v$ is defined similarly.

The two NMUFs v, $\tilde{v}$ on Y are said to be *cardinally equivalent* if there exist real numbers $\rho > 0$ and α such that $\tilde{v}(y) \equiv \alpha + \rho\, v(y)$ on Y. Then, as is well known, there is a unique cardinal equivalence class of NMUFs whose expected values all represent the ordering R on $\Delta(Y)$.

5.2. Continuous Behaviour

Rather than assume directly continuity of preferences on the domain of ordinary probabilities $\Delta(Y)$, however, it is in the spirit of consequentialist decision theory to postulate continuity of behaviour instead, whether or not that behaviour is in fact consequentialist. Accordingly, let T be any ordinary decision tree in $\mathcal{T}(Y)$, with strictly positive real probabilities at all its chance nodes. Let n^* be any decision node of T, and n^0 any chance node of the subtree $T(n^*) \in \mathcal{T}(Y)$ whose initial node is n^*. Consider now a family of ordinary decision trees $T^\pi \in \mathcal{T}(Y)$ in which only the probability distribution $\pi = \pi(\cdot|n^0) \in \Delta(N_{+1}(n^0))$ at the chance node n^0 varies. To ensure that $\pi(n) > 0$ for all $n \in N_{+1}(n^0)$ and so that $T^\pi \in \mathcal{T}(Y)$, remove from T the entire subtree $T(n)$ following any node $n \in N_{+1}(n^0)$ with $\pi(n) = 0$. All the other features of the decision trees T^π should be entirely independent of the probabilities π.

Now, the behaviour set $\beta(T^\pi, n^*)$ at the fixed decision node $n^* \in N^*$ is well defined and varies with π. Accordingly, one obtains a correspondence $\pi \mapsto\!\!\!\rightarrow \beta(T^\pi, n^*)$ whose graph is

$$G_\beta(T, n^*, n^0) := \{\, (\pi, n) \in \Delta(N_{+1}(n^0)) \times N_{+1}(n^*) \mid n \in \beta(T^\pi, n^*) \,\}.$$

Since $\Delta(N_{+1}(n^0))$ is compact while $N_{+1}(n^*)$ is a finite set, this correspondence will have a compact graph, and so be upper hemi-continuous, provided that its graph is a closed set. So behaviour β is said to be *continuous* provided that $G_\beta(T, n^*, n^0)$ is indeed closed, for every decision tree $T \in \mathcal{T}(Y)$, decision node $n \in N^*(T)$, and chance node $n^0 \in N^0(T(n^*))$. Formally, this is weaker than the stronger condition which requires β to be jointly continuous as *all* probabilities vary in each decision tree. In combination with the other axioms, however, the weaker continuity condition used here is actually equivalent to joint continuity.

The proof given in Hammond (1988b, Section 9) shows how such continuous behaviour implies the restricted continuity axiom stated in Section 5.1. Then there must indeed exist a unique cardinal equivalence class of von Neumann–Morgenstern utility functions whose expected values are maximized by the random consequences of behaviour in the restricted trees of $\mathcal{T}(Y)$.

6. Non-Archimedean Expected Utility

6.1. Representing Lexicographic Preferences

Suppose that non-Archimedean probabilities are RPFs taking values in the positive cone $\mathcal{P}(\epsilon)$. Suppose that the restricted continuity axiom of Section 5.1 is satisfied by consequentialist behaviour in the domain $\mathcal{T}(Y;\epsilon) := \mathcal{T}(Y;\mathcal{P}(\epsilon))$ of all finite decision trees with such RPFs attached to their chance nodes. Then the restriction of the ordering R to $\Delta(Y)$ is represented by the expected value of an NMUF $v : Y \to \Re$, as in (25).

Define the ordinary real-valued expected utility function $U : \Delta(Y) \to \Re$ so that

$$U(\lambda) := \mathbb{E}_\lambda\, v = \sum\nolimits_{y\in Y} \lambda(y)\, v(y) \tag{27}$$

for all $\lambda \in \Delta(Y)$. This definition can now be extended from ordinary probability distributions to RPFs in an obvious way. The result will be the *non-Archimedean expected utility function* $U(\cdot;\epsilon) : \Delta(Y;\epsilon) \to \Re(\epsilon)$ which is defined so that, given any

$$p = p(y;\epsilon) = \frac{\sum_{i\in I} P_i\, p_i(y)\, \epsilon^i}{\sum_{i\in I} P_i\, \epsilon^i} \in \Delta(Y;\epsilon),$$

one has

$$U(p;\epsilon) := \mathbb{E}_p\, v = \sum\nolimits_{y\in Y} p(y;\epsilon)\, v(y) = \frac{\sum_{i\in I} P_i \sum_{y\in Y} p_i(y)\, v(y)\, \epsilon^i}{\sum_{i\in I} P_i\, \epsilon^i} = \frac{\sum_{i\in I} P_i\, U(p_i)\, \epsilon^i}{\sum_{i\in I} P_i\, \epsilon^i}. \tag{28}$$

Suppose that the refinement axiom of Section 4.1 is imposed in addition. Section 4.2 characterized the resulting lexicographic preferences. Now these preferences can be interpreted as maximizing expected utility with respect to the total ordering on the space $\Re(\epsilon)$. To see this, let $p, q \in \Delta(Y;\epsilon)$ be any pair of rational probability functions of ϵ taking the form (20). They can be given a common denominator and put in the form

$$p = \frac{\sum_{k\in K} R_k\, p_k^*\, \epsilon^k}{\sum_{k\in K} R_k\, \epsilon^k} \quad \text{and} \quad q = \frac{\sum_{k\in K} R_k\, q_k^*\, \epsilon^k}{\sum_{k\in K} R_k\, \epsilon^k}$$

as in (23) of Section 4.2, with the ordinary probability distributions $p_k^*(\cdot), q_k^*(\cdot) \in \Delta(Y)$ given by (22), and with $R_k > 0$ (all $k \in K$). So $U(p;\epsilon) > U(q;\epsilon)$ in $\Re(\epsilon)$ if and only if

$$\frac{\sum_{k\in K} R_k\, U(p_k^*)\, \epsilon^k}{\sum_{k\in K} R_k\, \epsilon^k} - \frac{\sum_{k\in K} R_k\, U(q_k^*)\, \epsilon^k}{\sum_{k\in K} R_k\, \epsilon^k} = \frac{\sum_{k\in K} R_k\, [U(p_k^*) - U(q_k^*)]\, \epsilon^k}{\sum_{k\in K} R_k\, \epsilon^k} > 0 \text{ in } \Re(\epsilon),$$

or if and only if

$$\sum\nolimits_{k\in K} R_k \left[U(p_k^*) - U(q_k^*)\right] \epsilon^k > 0 \text{ in } \Re(\epsilon). \tag{29}$$

But (29) is true if and only if the following lexicographic criterion is satisfied: there must exist an integer $m \in K$ for which $U(p_k^*) = U(q_k^*)$ whenever $k \in K$ with $k < m$, while $U(p_m^*) > U(q_m^*)$. However, for all $k \in K$ one has $U(p_k^*) > U(q_k^*) \iff p_k^* \; P \; q_k^*$. Also, the theorem stated at the end of Section 4.2 shows that

$$p \; P \; q \iff \exists m \in K : p_k^* \; I \; q_k^* \text{ (all } k < m) \;\&\; p_m^* \; P \; q_m^*.$$

Finally, therefore, this chain of equivalences shows that $U(p;\epsilon) > U(q;\epsilon)$ in $\Re(\epsilon)$ if and only if $p \; P \; q$. This is true for all pairs $p, q \in \Delta(Y;\epsilon)$. So the preference ordering R is perfectly represented by the non-Archimedean expected utility function $U(\cdot;\epsilon)$, provided that the (lexicographic) total ordering in $\Re(\epsilon)$ is applied to each pair of non-Archimedean $\Re(\epsilon)$-valued expected utilities.

6.2. Why Refinement is Needed

LEMMA. *Suppose that $f(\epsilon), g(\epsilon) \in \mathcal{P}(\epsilon)$ with $f(\epsilon) > g(\epsilon)$. Suppose too that $p, q \in \Delta(Y;\epsilon)$ with $p \; P \; q$. Then, given any other $h(\epsilon) \in \mathcal{P}(\epsilon)$ and the two members $p_f, p_g \in \Delta(Y;\epsilon)$ defined by*

$$p_f := \frac{f(\epsilon)\,p + h(\epsilon)\,q}{f(\epsilon) + h(\epsilon)} \quad \text{and} \quad p_g := \frac{g(\epsilon)\,p + h(\epsilon)\,q}{g(\epsilon) + h(\epsilon)},$$

it must be true that $p_f \; P \; p_g$.

PROOF: For the non-Archimedean expected utility function $U(\cdot;\epsilon)$ which represents R on $\Delta(Y;\epsilon)$, one has

$$\begin{aligned} U(p_f) - U(p_g) &= \frac{f(\epsilon)\,U(p;\epsilon) + h(\epsilon)\,U(q;\epsilon)}{f(\epsilon) + h(\epsilon)} - \frac{g(\epsilon)\,U(p;\epsilon) + h(\epsilon)\,U(q;\epsilon)}{g(\epsilon) + h(\epsilon)} \\ &= \frac{[f(\epsilon) - g(\epsilon)]\,h(\epsilon)}{[f(\epsilon) + h(\epsilon)]\;[g(\epsilon) + h(\epsilon)]}\;[U(p;\epsilon) - U(q;\epsilon)] > 0 \end{aligned}$$

when $f(\epsilon) > g(\epsilon)$ and $p \; P \; q$, because then $f(\epsilon) - g(\epsilon)$, $f(\epsilon)$, $g(\epsilon)$, $h(\epsilon)$, and $U(p;\epsilon) - U(q;\epsilon)$ are all positive elements of $\Re(\epsilon)$. ∎

This lemma is important because it shows the need to distinguish the two members $f(\epsilon), g(\epsilon) \in \mathcal{P}(\epsilon)$ whenever $f(\epsilon) \neq g(\epsilon)$. For if these two were not distinguished, the corresponding two members $p_f, p_g \in \Delta(Y;\epsilon)$ would have to be regarded as identical, and so indifferent according to the conditional relation I. But then, since the order $\geq$ on $\Re(\epsilon)$ is total, one of the pair $f(\epsilon)$, $g(\epsilon)$ is greater than the other — say, $f(\epsilon) > g(\epsilon)$. Finally, the Lemma would imply that $p \; I \; q$ for all $p, q \in \Delta(Y; \epsilon)$, so that all random consequences would have to be indifferent. For this reason, non-trivial consequentialist decision theory requires all the different members of $\mathcal{P}(\epsilon)$ to be distinguished.

7. Main Theorem

Suppose probabilities take strongly positive values in the particular non-Archimedean field $\Re(\epsilon)$. For this particular case, the five assumptions introduced in previous Sections are:

A1) **Unrestricted Domain:** There is a non-empty behaviour set $\beta(T, n) \subset N_{+1}(n)$ at every decision node n of every tree T in the domain $\mathcal{T}(Y; \epsilon)$ of finite decision trees having (strongly positive) $\mathcal{P}(\epsilon)$-valued probabilities at all chance nodes and consequences in Y at all terminal nodes.

A2) **Dynamic Consistency:** Whenever $T \in \mathcal{T}(Y; \epsilon)$ and n' is a decision node of the subtree $T(n)$, then $\beta(T(n), n') = \beta(T, n')$.

A3) **Consequentialist Choice:** There is a revealed consequence choice function C_β such that the set $\Phi_\beta(T)$ of possible random consequences of behaviour β in any tree $T \in \mathcal{T}(Y; \epsilon)$ satisfies $\Phi_\beta(T) = C_\beta(F(T))$, where $F(T)$ is the set of random consequences which are feasible in T.

A4) **Refinement:** Given any tree $T \in \mathcal{T}(Y; \epsilon)$, let ${}^0T \in \mathcal{T}(Y)$ be the tree which is derived from T by replacing the $\mathcal{P}(\epsilon)$-valued probabilities $\pi(n'|n; \epsilon)$ $(n' \in N_{+1}(n))$ at each chance node n of T by their corresponding real parts ${}^0\pi(n'|n)$, followed by omitting any nodes which can only be reached with probability zero according to the probabilities ${}^0\pi$. Then one should have $\beta(T, n) \subset \beta({}^0T, n)$ at every decision node n of ${}^0T$.

A5) **Restricted Continuity:** For every ordinary decision tree $T \in \mathcal{T}(Y)$, decision node n^* in T, and chance node n^0 in the subtree $T(n^*)$ with initial node n^*, the correspondence $\pi \mapsto\!\!\!\rightarrow \beta(T^\pi, n^*)$ from real-valued probabilities $\pi \in \Delta(N_{+1}(n^0))$ to behaviour at n^* in the tree $T^\pi \in \mathcal{T}(Y)$ with probabilities π at n^0 has a closed graph.

Now:

Main Theorem.

(1a) Any behaviour satisfying the first four axioms A1)–A4) above reveals a (complete and transitive) preference ordering R defined on $\Delta(Y;\epsilon)$ with the properties that:

(i) There is a restricted preference ordering ${}^0R$ on the space $\Delta(Y)$ which, whenever $\lambda, \mu, \nu \in \Delta(Y)$ and $0 < \alpha < 1$, satisfies the independence condition

$$[\alpha\,\lambda + (1-\alpha)\,\nu]\ {}^0R\ [\alpha\,\mu + (1-\alpha)\,\nu] \iff \lambda\ {}^0R\ \mu.$$

(ii) Suppose that

$$p = \frac{\sum_{k\in K} R_k\, p_k^*\,\epsilon^k}{\sum_{k\in K} R_k\,\epsilon^k} \quad \text{and} \quad q = \frac{\sum_{k\in K} R_k\, q_k^*\,\epsilon^k}{\sum_{k\in K} R_k\,\epsilon^k}$$

are any two RPFs in $\Delta(Y;\epsilon)$ that have been given a common denominator, while $R_k > 0$ and $p_k^*, q_k^* \in \Delta(Y)$ for all $k \in K$, where K is a non-empty set of non-negative integers. Then one has $p\ P\ q$ if and only if the lexicographic criterion

$$\exists m \in K : p_k^*\ {}^0I\ q_k^*\ (k < m\ \&\ k \in K) \quad \text{and} \quad p_m^*\ {}^0P\ q_m^*$$

is satisfied.

(1b) Conversely, if R is any ordering defined on $\Delta(Y;\epsilon)$ that is derived by applying the lexicographic criterion in part (ii) of (1a) to an ordering ${}^0R$ on $\Delta(Y)$ satisfying the independence condition in part (i) of (1a), then behaviour that maximizes R in each decision tree of $\mathcal{T}(Y;\epsilon)$ will satisfy A1)–A4).

(2a) Behaviour will satisfy all five axioms A1)–A5) above only if, in addition, there is a unique cardinal equivalent class of NMUFs $v : Y \to \Re$ for which

$$\lambda\ {}^0R\ \mu \iff \sum_{y\in Y} [\lambda(y) - \mu(y)]\, v(y) \geq 0 \text{ whenever } \lambda, \mu \in \Delta(Y),$$

in which case the ordering R on $\Delta(Y;\epsilon)$ is the unique ordering induced by applying the (lexicographic) total order $\geq$ in $\Re(\epsilon)$ to values $U(p;\epsilon)$ of the non-Archimedean expected utility function that is defined for all $p \in \Delta(Y;\epsilon)$ by

$$U(p;\epsilon) := \sum_{y\in Y} p(y;\epsilon)\, v(y).$$

(2b) Conversely, maximizing the non-Archimedean expected value of any such NMUF in the (lexicographic) order of $\Re(\epsilon)$ will generate behaviour satisfying A1)–A5).

Proof: First, (1a) is proved by combining the arguments of Hammond (1988a and b, Sections 5 and 6) — see also Section 3.5 above — with the Theorem of Section 4.2. Conversely, (1b) is proved by a dynamic programming argument like that in Hammond (1988b, Section 8); the refinement axiom A4) is then obviously satisfied because of the lexicographic criterion in part (ii) of (1a).

For (2a), Section 5 explained why A5) would imply the existence of a unique cardinal equivalence class of NMUFs, and then Section 6.1 showed that behaviour would maximize non-Archimedean expected utility in the (lexicographic) total order $\geq$ of $\Re(\epsilon)$. Conversely, the single hypothesis of (2b) obviously implies all the hypotheses of (1b), so A1)–A4) will be satisfied. Furthermore, behaviour in ordinary decision trees with random consequences in $\Delta(Y)$ will maximize ordinary expected utility and so satisfy the restricted continuity axiom A5). ▮

An important corollary of this Main Theorem is that all attempts to treat as equivalent any pair of non-Archimedean probability distributions having different values in $\mathcal{P}(\epsilon)$ are doomed to fail. For, when combined with the five axioms set out above, any such attempt leads to the unacceptable conclusion that all non-Archimedean random consequences are indifferent — see Section 6.2.

ACKNOWLEDGEMENTS

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 entitled "Consequentialism, Extended Probabilities, and Lexicographic Expected Utility" was written during a fruitful visit, arranged by Joe Stiglitz, to the Department of Economics at Princeton University in the Spring semester of 1987. Later, in Spring 1988 I received support from the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 303 while visiting the University of Bonn as a guest of Dieter Bös.

The work reported here and in related papers 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 Reinhard Selten's 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, as well as the University of California at San Diego. Other improvements and encouragement have resulted from helpful comments by several seminar participants, Bill Zame, Reinhard Selten, Barry Nalebuff, Avinash Dixit, Andrew Caplin, and Pierpaolo Battigalli in particular. My thanks to all the above individuals and institutions.

REFERENCES

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.

HAMMOND, P.J. (1976), "Changing Tastes and Coherent Dynamic Choice," *Review of Economic Studies*, **43**: 159–173.

HAMMOND, P.J. (1988a), "Consequentialism and the Independence Axiom," in *Risk, Decision, and Rationality: Proceedings of the 3rd International Conference on the Foundations and Applications of Utility, Risk, and Decision Theories* edited by B. Munier (Dordrecht: D. Reidel), pp. 503–516.

HAMMOND, P.J. (1988b), "Consequentialist Foundations for Expected Utility," *Theory and Decision*, **25**: 25–78.

HAMMOND, P.J. (1994), "Elementary Non-Archimedean Representations of Probability for Decision Theory and Games," in *Patrick Suppes: Scientific Philosopher, Vol. I: Probability and Probabilistic Causality* edited by P. Humphreys (Dordrecht: Kluwer Academic Publishers), ch. 2, pp. 25–59.

HERSTEIN, I.N. AND J. MILNOR (1953), "An Axiomatic Approach to Measurable Utility," *Econometrica*, **21**: 291–297.

JENSEN, N.E. (1967), "An Introduction to Bernoullian Utility Theory, I: Utility Functions," *Swedish Journal of Economics*, **69**: 163–183.

KREPS, D. AND R. WILSON (1982), "Sequential Equilibrium," *Econometrica*, **50**: 863–894.

LINDLEY, D.V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 1: Probability*. Cambridge: Cambridge University Press.

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.

Raiffa, H. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, Mass.: Addison-Wesley.

Rényi, A. (1955), "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**: 285–335.

Rényi, A. (1970), *Probability Theory*. New York: Elsevier.

Robinson, A. (1973), "Function Theory on Some Nonarchimedean Fields," *American Mathematical Monthly: Papers in the Foundations of Mathematics*, **80**: S87–S109.

Samuelson, P.A. (1952), "Probability, Utility, and the Independence Axiom," *Econometrica*, **20**: 670–678.

Selten, R. (1975), "Re-examination of the Perfectness Concept for Equilibrium Points of Extensive Games," *International Journal of Game Theory*, **4**: 25–55.

von Neumann, J. and O. Morgenstern (1944, 1953), *Theory of Games and Economic Behavior (3rd. edn.)*. Princeton: Princeton University Press.