Lexicographic probability, conditional probability, and nonstandard probability*

Joseph Y. Halpern
Dept. Computer Science
Cornell University
Ithaca, NY 14853
halpern@cs.cornell.edu
http://www.cs.cornell.edu/home/halpern

June 17, 2003

Abstract

The relationship between *Popper spaces* (conditional probability spaces that satisfy some regularity conditions), lexicographic probability systems (LPS's) [Blume, Brandenburger, and Dekel 1991a; Blume, Brandenburger, and Dekel 1991b], and nonstandard probability spaces (NPS's) is considered. If countable additivity is assumed, Popper spaces and a subclass of LPS's are equivalent; without the assumption of countable additivity, the equivalence no longer holds. If the state space is finite, LPS's are equivalent to NPS's. However, if the state space is infinite, NPS's are shown to be more general than LPS's.

1 Introduction

Probability is certainly the most commonly-used approach for representing uncertainty and conditioning the standard way of updating probabilities in the light of new information. Unfortunately, there is a well-known problem with conditioning: Conditioning on events of measure 0 is not defined. That makes it unclear how to proceed if an agent learns something to which she initially assigned probability 0. Although conditioning on events of measure 0 may seem to be of little practical interest, it turns out to play a critical role in game theory (see, for example, [Blume, Brandenburger, and Dekel 1991a; Blume, Brandenburger, and Dekel 1991b; Hammond 1994; Kreps and Wilson 1982; Myerson 1986; Selten 1965]), the analysis of conditional statements (see [Adams 1966; McGee 1994]), and in dealing with nonmonotonicity (see, for example, [Lehmann and Magidor 1992]).

There have been various attempts to deal with the problem of conditioning on events of measure 0. Perhaps the best known, which goes back to Popper [1968] and de Finetti [1936], is

*The work was supported in part by NSF under grants IRI-96-25901, IIS-0090145, and CTC-0208535, by ONR under grant N00014-02-1-0455, and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grant N00014-01-1-0795.

to take as primitive not probability, but conditional probability. If μ is a conditional probability measure, then $\mu(V \mid U)$ may still be undefined for some pairs V and U, but it is also possible that $\mu(V \mid U)$ is defined even if $\mu(U) = 0$. Another approach, which goes back to at least Robinson [1973] and has been explored in the economics literature [Hammond 1994], the AI literature [Lehmann and Magidor 1992; Wilson 1995], and the philosophy literature (see [McGee 1994] and the references therein) is to consider *nonstandard probability spaces* (NPS's), where there are infinitesimals that can be used to model events that, intuitively, have infinitesimally small probability yet may still be learned or observed.

There is another approach to this problem, which uses sequences of probability measures to represent uncertainty. The most recent exemplar of this approach, which I focus on here, are the *lexicographic probability systems* of Blume, Brandenburger, and Dekel [1991a, 1991b] (BBD from now on). However, the idea of using a system of measures to represent uncertainty actually was explored as far back as the 1950s by Rényi [1956] (see Section 3.3). A *lexicographic probability system* is a sequence $\langle \mu_0, \mu_1, \ldots \rangle$ of probability measures. Intuitively, the first measure in the sequence, μ_0, is the most important one, followed by μ_1, μ_2, and so on. Roughly speaking, the probability assigned to an event U by a sequence such as $\langle \mu_0, \mu_1 \rangle$ can be taken to be $\mu_0(U) + \epsilon\mu_1(U)$, where ϵ is an infinitesimal. Thus, even if the probability of U according to μ_0 is 0, U still has a positive (although infinitesimal) probability if $\mu_1(U) > 0$.

How are all these approaches related? This question, which is the focus of the paper, has been considered before. For example, Hammond [1994] shows that conditional probability spaces are equivalent to a subclass of LPS's called lexicographic conditional probability spaces if the state space is finite and it is possible to condition on any nonempty set. As shown by Spohn [1986], Hammond's result can be extended to arbitrary countably additive *Popper spaces*, where a Popper space is a conditional probability space that satisfies certain regularity conditions. The extension is nontrivial and, indeed, does not work without the assumption of countable additivity. Rényi [1956] and van Fraassen [1976] provide other representations of conditional probability spaces as sequences of measures, although not LPS's. Their results apply even if the underlying state space is infinite, but countable additivity does not play a role in their representations. (See Section 3 for further discussion of this issue.)

I show that if the state space is finite, then LPS's are equivalent to NPS's, using a strong notion of equivalence. This equivalence breaks down if the state space is infinite; in this case, NPS's are strictly more general than LPS's (whether or not countable additivity is assumed).

Finally, I consider the relationship between Popper spaces and NPS's, and show that NPS's are more general. (The theorem I prove is a generalization of one proved by McGee [1994], but my interpretation of it is quite different; see Section 5.)

The remainder of the paper is organized as follows. In the next section, I review all the relevant definitions for the three representations of uncertainty considered here. Section 3 considers the relationship between Popper spaces and LPS's. Section 4 considers the relationship between LPS's and NPS's. Finally, Section 5 considers the relationship between Popper spaces and NPS's. In Section 6 I consider what these results have to say about independence. I conclude with some discussion in Section 7.

2 Conditional, lexicographic, and nonstandard probability spaces

In this section I briefly review the three approaches to representing likelihood discussed in the introduction.

2.1 Popper spaces

A *conditional probability measure* takes *pairs* U, V of subsets as arguments; $\mu(V, U)$ is generally written $\mu(V \mid U)$ to stress the conditioning aspects. The first argument comes from some algebra $\mathcal{F}$ of subsets of a space W; if W is infinite, $\mathcal{F}$ is often taken to be a σ-algebra. (Recall that an algebra of subsets of W is a set of subsets containing W and closed under union and complementation. A σ-algebra is an algebra that is closed under union countable.) The question is what constraints, if any, should be placed on the second argument. I start with three minimal requirements, and later add a fourth.

Definition 2.1: A *Popper algebra* over W is a set $\mathcal{F} \times \mathcal{F}'$ of subsets of $W \times W$ such that (a) $\mathcal{F}$ is an algebra over W, (b) $\mathcal{F}'$ is a nonempty subset of $\mathcal{F}$ (not necessarily an algebra over W), and (c) $\mathcal{F}'$ is closed under supersets in $\mathcal{F}$, in that if $V \in \mathcal{F}'$, $V \subseteq V'$, and $V' \in \mathcal{F}$, then $V' \in \mathcal{F}'$. (Popper algebras are named after Karl Popper.) ∎

Definition 2.2: A *conditional probability space (cps) over* $(W, \mathcal{F})$ is a tuple $(W, \mathcal{F}, \mathcal{F}', \mu)$ such that $\mathcal{F} \times \mathcal{F}'$ is a Popper algebra over W and $\mu : \mathcal{F} \times \mathcal{F}' \to [0, 1]$ satisfies the following conditions:

CP1. $\mu(U \mid U) = 1$ if $U \in \mathcal{F}'$.

CP2. $\mu(V_1 \cup V_2 \mid U) = \mu(V_1 \mid U) + \mu(V_2 \mid U)$ if $V_1 \cap V_2 = \emptyset$, $U \in \mathcal{F}'$, and $V_1, V_2 \in \mathcal{F}$.

CP3. $\mu(V \mid U) = \mu(V \mid X) \times \mu(X \mid U)$ if $V \subseteq X \subseteq U$, $U, X \in \mathcal{F}'$, $V \in \mathcal{F}$.

A *Popper space over* $(W, \mathcal{F})$ is a conditional probability space $(W, \mathcal{F}, \mathcal{F}', \mu)$ that satisfies an additional condition: if $U \in \mathcal{F}'$ and $\mu(V \mid U) \neq 0$ then $V \cap U \in \mathcal{F}'$. If $\mathcal{F}$ is a σ-algebra and μ is countably additive (that is, if $\mu(\cup V_i \mid U) = \sum_{i=1}^{\infty} \mu(V_i \mid U)$ if the V_i's are pairwise disjoint elements of $\mathcal{F}$ and $U \in \mathcal{F}'$), then the Popper space is said to be *countably additive*. Let $Pop(W, \mathcal{F})$ denote the set of Popper spaces over $(W, \mathcal{F})$; if $\mathcal{F}$ is a σ-algebra, let $Pop^c(W, \mathcal{F})$ denote the set of countably additive Popper spaces over $(W, \mathcal{F})$. The probability measure μ in a Popper space is called a *Popper measure*. ∎

The additional regularity condition on $\mathcal{F}'$ required in a Popper space corresponds to the observation that for an unconditional probability measure μ, if $\mu(V \mid U) \neq 0$ then $\mu(V \cap U) \neq 0$, so conditioning on $V \cap U$ should be defined.

Popper [1968] was the first to consider formally conditional probability as the basic notion, although his definition of conditional probability space is not quite the same as that used here. CP1–3 are essentially due to Rényi [1955]. De Finetti [1936] also did some early work, apparently independently, taking conditional probabilities as primitive. Indeed, as Rényi [1964] points out, the idea of taking conditional probability as primitive seems to go back as far as Keynes [1921]. Van Fraassen [1976] defined what I have called Popper measures; he called them Popper functions, reserving the name Popper measure for what I am calling a countably additive Popper measure. Hammond [1994] discusses the use of conditional probability spaces in philosophy and game theory, and provides an extensive list of references.

2.2 Lexicographic probability spaces

Definition 2.3: A *lexicographic probability space (LPS) (of length α) over* $(W, \mathcal{F})$ is a tuple $(W, \mathcal{F}, \vec{\mu})$ where, as before, W is a set of possible worlds and $\mathcal{F}$ is an algebra over W, and $\vec{\mu}$ is a sequence of probability measures on $(W, \mathcal{F})$ indexed by ordinals $< \alpha$. (Technically, $\vec{\mu}$ is a function from the ordinals less than α to probability measures on $(W, \mathcal{F})$.) I typically write $\vec{\mu}$ as $(\mu_0, \mu_1, \ldots)$ or as $(\mu_\beta : \beta < \alpha)$. If $\mathcal{F}$ is a σ-algebra and each of the probability measures in $\vec{\mu}$ is countably additive, then $\vec{\mu}$ is a *countably additive LPS*. Let $LPS(W, \mathcal{F})$ denote the set of LPS's over $(W, \mathcal{F})$; if $\mathcal{F}$ is a σ-algebra, let $LPS^c(W, \mathcal{F})$ denote the set of countably additive LPS's over $(W, \mathcal{F})$. When $(W, \mathcal{F})$ are understood, I often refer to $\vec{\mu}$ as the LPS. ∎

BBD define a *lexicographic conditional probability space (LCPS)* to be an LPS such that the probability measures in the sequence have disjoint supports; that is, there exist sets $U_i \in \mathcal{F}$ such that $\mu_i(U_i) = 1$ and the sets U_i are pairwise disjoint for $i < \alpha$. Let a *structured LPS (SLPS)* be an LPS such that there exist sets $U_i \in \mathcal{F}$ such that $\mu_i(U_i) = 1$ and $\mu_i(U_j) = 0$ for $j > i$. (Spohn [1986] calls SLPS's *dimensionally well-ordered families of probability measures*.) Let $SLPS(W, \mathcal{F})$ denote the set of SLPS's over $(W, \mathcal{F})$; if $\mathcal{F}$ is a σ-algebra, let $SLPS^c(W, \mathcal{F})$ denote the set of countably additive SLPS's over $(W, \mathcal{F})$.

Clearly every LCPS is an SLPS. Moreover, if α is countable, then every countably additive SLPS is an LCPS: Given an SLPS $\vec{\mu}$ with associated sets $U_i, i < \alpha$, define $U_i' = U_i - (\cup_{j>i} U_j)$. The sets U_i' are clearly pairwise disjoint elements of $\mathcal{F}$, and U_i' is a support for μ_i. Of course, the same argument holds even without the assumption of countable additivity if α is finite. However, in general, LCPS's are a strict subset of SLPS's, as the following example shows.

Example 2.4: Consider a well-ordering of the interval $[0, 1]$, that is, an isomorphism from $[0, 1]$ to an initial segment of the ordinals. Suppose that this initial segment of the ordinals has length α. Let $([0, 1], \mathcal{F}, \vec{\mu})$ be an LPS of length α where $\mathcal{F}$ consists of the Borel subsets of $[0, 1]$. Let μ_0 be the standard Borel measure on $[0, 1]$, and let μ_β be the measure that gives probability 1 to r_β, the βth real in the well-ordering. This clearly gives an SLPS, since the support of μ_0 is $[0, 1]$ and the support of μ_β for $0 < \beta < \alpha$ is $\{r_\beta\}$. However, this SLPS is not equivalent to any LCPS; there is no support of μ_0 which is disjoint from the supports of μ_β for all β with $0 < \beta < \alpha$. ∎

The difference between LCPS's and SLPS's does not arise in the work of BBD, since they consider only finite sequences of measures. The restriction to finite sequences, in turn, is due to their restriction to finite sets W of possible worlds. Clearly, if W is finite, then all LCPS's over W must have length $\leq |W|$, since the measures in an LCPS have disjoint supports.

We can put an obvious lexicographic order $<_L$ on sequences $(x_0, x_1, \ldots)$ of numbers in $[0, 1]$ of length α: $(x_0, x_1, \ldots) <_L (y_0, y_1, \ldots)$ if there exists $\beta < \alpha$ such that $x_\beta < y_\beta$ and $x_\gamma = y_\gamma$ for all $\gamma < \beta$. That is, we compare two sequences by comparing their components at the first place they differ. (Even if α is infinite, because we are dealing with ordinals, there will be a least ordinal at which the sequences differ if they differ at all.) This lexicographic order will be used to define decision rules.

BBD define conditioning in LPS's as follows. Given $\vec{\mu}$ and $U \in \mathcal{F}$ such that $\mu_i(U) > 0$ for some index i, let $\vec{\mu}|U = (\mu_{k_0}(\cdot \mid U), \mu_{k_1}(\cdot \mid U), \ldots)$, where $(k_0, k_1, \ldots)$ is the subsequence of all indices for which the probability of U is positive. Formally, $k_0 = \min\{k : \mu_k(U) > 0\}$ and for an arbitrary ordinal $\beta > 0$, if μ_{k_γ} has been defined for all $\gamma < \beta$ and there exists a measure μ_δ in $\vec{\mu}$ such that $\mu_\delta(U) > 0$ and $\delta > k_\gamma$ for all $\gamma < \beta$, then $k_\beta = \min\{\delta : \mu_\delta(U) > 0, \delta > k_\gamma \text{ for all } \gamma < \beta\}$. Note that $\vec{\mu}|U$ is undefined if $\mu_\beta(U) = 0$ for all $\beta < \alpha$.

2.3 Nonstandard probability spaces

It is well known that there exist *non-Archimedean fields*—fields that include the real numbers as a subfield but also have *infinitesimals*, numbers that are positive but still less than any positive real number. The smallest such non-Archimedean field, commonly denoted $I\!R(\epsilon)$, is the smallest field generated by adding to the reals a single infinitesimal ϵ.[1] The *hyperreals*, nonstandard models of the reals that satisfy all the first-order properties that hold of the real numbers (see [Davis 1977]), are also instances of non-Archimedean fields. For most of this paper, I use only the following properties of non-Archimedean fields:

1. If $I\!R^*$ is a non-Archimedean field, then for all $b \in I\!R^*$ such that $-r < b < r$ for some standard real $r > 0$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. (Formally, a is the inf of the set of real numbers that are at least as large as b.) Let $st(b)$ denote the closest standard real to b; $st(b)$ is sometimes read "the standard part of b".

2. If $st(\epsilon/\epsilon') = 0$, then $a\epsilon < \epsilon'$ for all positive standard real numbers a. (If $a\epsilon$ were greater than ϵ', then ϵ/ϵ' would be greater than $1/a$, contradicting the assumption that $st(\epsilon/\epsilon') = 0$.)

Given a non-Archimedean field $I\!R^*$, a *nonstandard probability space (NPS) over* $(W, \mathcal{F})$ *(with range* $I\!R^*$*)* is a tuple $(W, \mathcal{F}, \mu)$, where W is a set of possible worlds, $\mathcal{F}$ is an algebra of subsets of W, and μ assigns to sets in $\mathcal{F}$ an element of $I\!R^*$ such that $\mu(W) = 1$ and $\mu(U \cup V) = \mu(U) + \mu(V)$ if U and V are disjoint. If W is infinite, we may also require that $\mathcal{F}$ be a σ-algebra and that μ be countably additive. (There are some subtleties involved with countable additivity in nonstandard probability spaces; see Section 4.3.)

3 Relating Popper Spaces to (S)LPS's

In this section, I consider a mapping $F_{S \to P}$ from SLPS's over $(W, \mathcal{F})$ to Popper spaces over $(W, \mathcal{F})$, for each fixed W and $\mathcal{F}$, and show that, in many cases of interest, $F_{S \to P}$ is an isomorphism. Given an SLPS $(W, \mathcal{F}, \vec{\mu})$ of length α, consider the cps $(W, \mathcal{F}, \mathcal{F}', \mu)$ such that $\mathcal{F}' = \cup_{\beta < \alpha}\{V \in \mathcal{F} : \mu_\beta(V) > 0\}$. For $V \in F'$, let j_V be the smallest index such $\mu_{j_V}(V) > 0$.

[1] The construction of $I\!R(\epsilon)$ apparently goes back to Robinson [1973]. It is reviewed by Hammond [1994] and Wilson [1995] (who calls $I\!R(\epsilon)$ the *extended reals*).

Define $\mu(U \mid V) = \mu_{j_V}(U \mid V)$. I leave it to the reader to check that $(W, \mathcal{F}, \mathcal{F}', \mu)$ is a Popper space.

There are many isomorphisms between two spaces. Why is $F_{S \to P}$ of interest? Suppose that $F_{S \to P}(W, \mathcal{F}, \vec{\mu}) = (W, \mathcal{F}, \mathcal{F}', \mu)$. It is easy to check that the following two important properties hold:

- $\mathcal{F}'$ consists precisely of those events for which conditioning in the LPS is defined; that is, $\mathcal{F}' = \{U : \mu_\beta(U) \neq 0 \text{ for some } \mu_\beta \in \vec{\mu}\}$.
- For $U \in \mathcal{F}'$, $\mu(\cdot \mid U) = \mu'(\cdot \mid U)$, where μ' is the first probability measure in the sequence $\vec{\mu}|U$. That is, the Popper measure agrees with the most significant probability measure in the conditional LPS given U. Given that an LPS assigns to an event U a sequence of numbers and a Popper measure assigns to U just a single number, this is clearly the best single number to take.

It seems that these are minimal properties that an isomorphism should satisfy. Moreover, it is easy to see that these two properties completely characterize $F_{S \to P}$.

3.1 The finite case

It is useful to separate the analysis of $F_{S \to P}$ into two cases, depending on whether or not the state space is finite. I consider the finite case first.

BBD claim without proof that $F_{S \to P}$ is an isomorphism from LCPS's to conditional probability spaces. They work in finite spaces W (so that LCPS's are equivalent to SLPS's) and restrict attention to LPS's where $\mathcal{F} = 2^W$ and $\mathcal{F}' = 2^W - \emptyset$ (so that conditioning is defined for all nonempty sets). Since $\mathcal{F}' = 2^W - \emptyset$, the cps's they consider are all Popper spaces. Hammond [1994] provides a careful proof of this result, under the restrictions considered by BBD. I generalize Hammond's result by considering arbitrary finite Popper spaces. No new conceptual issues arise in doing this extension; I include it here only to be able to contrast it with the other results.

Theorem 3.1: *If W is finite, then $F_{S \to P}$ is an isomorphism from* SLPS$(W, \mathcal{F})$ *to* Pop$(W, \mathcal{F})$.

3.2 The infinite case

The case where the state space W is infinite is not considered by either BBD or Hammond. It presents some interesting subtleties.

It is easy to see that $F_{S \to P}$ is an injection from from SLPS's to Popper spaces. However, as the following two examples show, if we do not require countable additivity, it is not an isomorphism.

Example 3.2: (This example is essentially due to Robert Stalnaker [private communication, 2000].) Let $W = I\!N$, the natural numbers, let $\mathcal{F}$ consist of the finite and cofinite subsets of $I\!N$, and let $\mathcal{F}' = \mathcal{F} - \{\emptyset\}$. If U is cofinite, take $\mu^1(V \mid U)$ to be 1 if V is cofinite and 0 if V is finite. If U is finite, define $\mu^1(V \mid U) = |V \cap U|/|U|$. I leave it to the reader to check that

$(\mathbb{N}, \mathcal{F}, \mathcal{F}', \mu^1)$ is a Popper space. Suppose there were some LPS $(\mathbb{N}, \mathcal{F}, \vec{\mu})$ which was mapped by $F_{S\to P}$ to this Popper space. Then it is easy to check that if μ_i is the first measure in $\vec{\mu}$ such that $\mu_i(U) > 0$ for some finite set U, then $\mu_i(U') > 0$ for all nonempty finite sets U'. To see this, note that for any nonempty finite set U', since $\mu_i(U) > 0$, it follows that $\mu_i(U \cup U') > 0$. Since $U \cup U'$ is finite, it must be the case that μ_i is the first measure in $\vec{\mu}$ such that $\mu_i(U \cup U') > 0$. Thus, by definition, $\mu^1(U' \mid U \cup U') = \mu_i(U' \mid U \cup U')$. Since $\mu^1(U' \mid U \cup U') > 0$, it follows that $\mu_i(U') > 0$. Thus, $\mu_i(U') > 0$ for all nonempty finite sets U'.

It is also easy to see that $\mu_i(U)$ must be proportional to $|U|$ for all finite sets U. To show this, it clearly suffices to show that $\mu_i(n) = \mu_i(0)$ for all $n \in \mathbb{N}$. But this is immediate from the observation that

$$\mu_i(\{0\} \mid \{0, n\}) = \mu^1(\{0\} \mid \{0, n\}) = |\{0\}|/|\{0, n\}| = \frac{1}{2}.$$

But there is no countably probability measure μ_i on the natural numbers that gives all natural numbers the same measure. For, by countable additivity, if $\mu_i(0) = 0$ then $\mu_i(\mathbb{N}) = 0$ and if $\mu_i(0) > 0$, then $\mu_i(\mathbb{N}) = \infty$. ∎

Example 3.3: Again, let $W = \mathbb{N}$, let $\mathcal{F}$ consist of the finite and cofinite subsets of $\mathbb{N}$, and let $\mathcal{F}' = \mathcal{F} - \{\emptyset\}$. As with μ^1, if U is cofinite, take $\mu^2(V \mid U)$ to be 1 if V is cofinite and 0 if V is finite. However, now, if U is finite, define $\mu^2(V \mid U) = 1$ if $\max(V \cap U) = \max V$, and $\mu^2(V \mid U) = 0$ otherwise. Intuitively, if $n > n'$, then n is infinitely more probable than n' according to μ^2. Again, I leave it to the reader to check that $(\mathbb{N}, \mathcal{F}, \mathcal{F}', \mu^2)$ is a Popper space. Suppose there were some LPS $(\mathbb{N}, \mathcal{F}, \vec{\mu})$ which was mapped by $F_{S\to P}$ to this Popper space. Then it is easy to check that if μ_n is the first measure in $\vec{\mu}$ such that $\mu_n(\{n\}) > 0$, then μ_n comes before $\mu_{n'}$ in $\vec{\mu}$ if $n > n'$. However, since $\vec{\mu}$ is well-founded, this is impossible. ∎

As the following theorem, proved by Spohn [1986], shows, there are no such counterexamples if we restrict to countably additive SLPS's and countably additive Popper spaces.

Theorem 3.4: [Spohn 1986] *For all W, the map $F_{S\to P}$ is an isomorphism from* $\mathrm{SLPS}^c(W, \mathcal{F})$ *to* $\mathrm{Pop}^c(W, \mathcal{F})$.

It is important in Theorem 3.4 that we consider SLPS's and not LCPS's. $F_{S\to P}$ is in fact not an isomorphism from LCPS's to Popper spaces.

Example 3.5: Consider the Popper space $([0, 1], \mathcal{F}, \mathcal{F}', \mu)$ which is the image under $F_{S\to P}$ of the SLPS constructed in Example 2.4. It is easy to see that this Popper space cannot be the image under $F_{S\to P}$ of some LCPS. ∎

3.3 Related Work

It is interesting to contrast these results to those of Rényi [1956] and van Fraassen [1976]. Renyi considers what he calls *dimensionally ordered* systems. A dimensionally ordered system over $(W, \mathcal{F})$ has the form $(W, \mathcal{F}, \mathcal{F}', \{\mu_i : i \in I\})$, where $\mathcal{F}$ is a an algebra of subsets of W, $\mathcal{F}'$ is a subset of $\mathcal{F}$ closed under finite unions, I is a totally ordered set (but not necessarily well-founded, so it may not, for example, have a first element) and μ_i is a measure on $(W, \mathcal{F})$ (not necessarily a probability measure) such that

- for each $U \in \mathcal{F}'$, there is some $i \in I$ such that $0 < \mu_i(U) < \infty$ (note that the measure of a set may, in general, be ∞),
- if $\mu_i(U) < \infty$ and $j < i$, then $\mu_j(U) = 0$.

Note that it follows from these conditions that for each $U \in \mathcal{F}'$, there is exactly one $i \in I$ such that $0 < \mu_i(U) < \infty$.

There is an obvious analogue of the map $F_{S\to P}$ mapping dimensionally ordered system to cps's. Namely, let $F_{D\to C}$ map the dimensionally ordered system $(W, \mathcal{F}, \mathcal{F}', \{\mu_i : i \in I\})$ to the cps $(W, \mathcal{F}, \mathcal{F}', \mu)$, where $\mu(V \mid U) = \mu_i(V \mid U)$, where i is the unique element of I such that $0 < \mu_i(U) < \infty$. Rényi shows that $F_{D\to C}$ is an isomorphism from dimensionally ordered systems to cps's where the set $\mathcal{F}'$ is closed under finite unions. (Csaszar [1955] extends this result to cases where the set $\mathcal{F}'$ is not necessarily closed under finite unions.) Rényi assumes that all measures involved are countably additive and that $\mathcal{F}$ is a σ-algebra, but these are inessential assumptions. That is, his proof goes through without change if $\mathcal{F}$ is an algebra and the measures are additive; all that happens is that the resulting conditional probability measure is additive rather than σ-additive.

It is critical in Rényi's framework that the μ_i's are arbitrary measures, and not just probability measures. His result does not hold if the μ_i's are required to be probability measures. In the case of finitely additive measures, the Popper space constructed in Example 3.2 already shows why. It corresponds to a dimensionally ordered space (μ_1, μ_2) where $\mu_1(U)$ is 1 if U is cofinite and 0 if U is finite and $\mu_2(U) = |U|$ (i.e., the measure of a set is its cardinality). It cannot be captured by a dimensionally ordered space where all the elements are probability measures, for the same reason that it is not the image of an SLPS under $F_{S\to P}$. (Rényi [1956] actually provides a general characterization of when the μ_i's can be taken to be (countably additive) probability measures.)

Another example is provided by the Popper space considered in Example 3.3. This corresponds to the dimensionally ordered system $\{\mu_\beta : \beta \in \mathbb{N} \cup \{\infty\}\}$, where

$$\mu_n(U) = \begin{cases} 0 & \text{if } \max(U) < n \\ 1 & \text{if } \max(U) = n \\ \infty & \text{if } \max(U) > n, \end{cases}$$

where $\max(U)$ is taken to be ∞ if U is cofinite.

Van Fraassen [1976] proved a result whose assumptions are somewhat closer to Theorem 3.4. Van Fraassen considers what he calls *ordinal families of probability measures*. An ordinal family over $(W, \mathcal{F})$ is a sequence of the form $\{(W_\beta, \mathcal{F}_\beta, \mu_\beta) : \beta < \alpha\}$ such that

- $\cup_{\beta<\alpha} W_\beta = W$;
- $\mathcal{F}_\beta$ is an algebra over W_β;
- μ_β is a probability measure with domain $\mathcal{F}_\beta$;
- $\cup_{\beta<\alpha} \mathcal{F}_\beta = \mathcal{F}$;
- if $U \in \mathcal{F}$ and $V \in \mathcal{F}_\beta$, then $U \cap V \in \mathcal{F}_\beta$;

- if $U \in \mathcal{F}$, $U \cap V \in \mathcal{F}_\beta$, and $\mu_\beta(U \cap V) > 0$, then there exists γ such that $U \in \mathcal{F}_\gamma$ and $\mu_\gamma(U) > 0$.

Given an ordinal family $\{(W_\beta, \mathcal{F}_\beta, \mu_\beta) : \beta < \alpha\}$ over $(W, \mathcal{F})$, consider the map $F_{O \rightarrow C}$ which associates with it the cps $(W, \mathcal{F}, \mathcal{F}', \mu)$, where $\mathcal{F}' = \{U \in \mathcal{F} : \mu_\gamma(U) > 0 \text{ for some } \gamma < \alpha\}$ and $\mu(V \mid U) = \mu_\beta(V \mid U)$, where β is the smallest ordinal such that $U \in \mathcal{F}_\beta$ and $\mu_\beta(U) > 0$. Van Fraassen shows that $F_{O \rightarrow C}$ is an isomorphism from ordinal families over $(W, \mathcal{F})$ to Popper spaces over $(W, \mathcal{F})$. Again, for van Fraassen, countable additivity does not play a significant role. If $\mathcal{F}$ is a σ-algebra, a *countably additive* ordinal family over $(W, \mathcal{F})$ is defined just as an ordinal family, except that now $\mathcal{F}_\beta$ is a σ-algebra over W_β for all $\beta < \alpha$, μ_α is a countably additive probability measure, and $\mathcal{F}$ is the least σ-algebra containing $\cup_{\beta<\alpha}\mathcal{F}_\beta$ (since $\cup_{\beta<\alpha}\mathcal{F}_\beta$ is not in general a σ-algebra). The same map $F_{O \rightarrow C}$ is also an isomorphism from countably additive ordinal families to countably additive Popper spaces.

Spohn's result, Theorem 3.4, can be viewed as a strengthening of van Fraassen's result in the countably additive case, since for Theorem 3.4 all the $\mathcal{F}_\beta$'s are required to be identical. This is a nontrivial requirement. The fact that it cannot be met in the case that W is infinite and the measures are not countably additive is an indication of this.

It is worth seeing how van Fraassen's approach handles the finitely additive examples which do not correspond to SLPS's. The Popper space in Example 3.2 corresponds to the ordinal family $\{(W_n, \mathcal{F}_n, \mu_n) : n \leq \omega\}$ where, for $n < \omega$, $W_n = \{1, \ldots, n\}$, $\mathcal{F}_n$ consists of all subsets of W_n, and μ_n is the uniform measure, while $W_\omega = I\!N$, $\mathcal{F}_\omega$ consists of the finite and cofinite subsets of $I\!N$, and $\mu_\omega(U)$ is 1 if U is cofinite and 0 if U is finite. It is easy to check that this ordinal family has the desired properties. The Popper space in Example 3.3 is represented in a similar way, using the ordinal family $\{(W_n, \mathcal{F}_n, \mu'_n) : n \leq \omega\}$, where $\mu'_n(U)$ is 1 if $n \in U$ and 0 otherwise, while $\mu'_\omega = \mu_\omega$. I leave it to the reader to see that this family has the desired properties. The key point to observe here is the leverage obtained by allowing each probability measure to have a different domain.

4 Relating LPS's to NPS's

In this section, I show that LPS's and NPS's are isomorphic in a strong sense. Again, I separate the results for the finite case and the infinite case.

4.1 The finite case

Consider an LPS of the form (μ_1, μ_2, μ_3). Roughly speaking, the corresponding NPS should be $(1 - \epsilon - \epsilon^2)\mu_1 + \epsilon\mu_2 + \epsilon^2\mu_3$, where ϵ is some infinitesimal. That means that μ_2 gets infinitesimal weight relative to μ_1 and μ_3 gets infinitesimal weight relative to μ_2. But which infinitesimal ϵ should be chosen? Intuitively, it shouldn't matter. No matter which infinitesimal is chosen, the resulting NPS should be equivalent to the original LPS. How can we make this intuition precise?

Suppose that we want to use an LPS or an NPS to compute which of two bounded, *real-valued* random variables has higher expected value. (The intended application here is decision making, where the functions can be thought of as the utilities corresponding to two actions;

the one with higher expected utility is preferred.) The idea is that two measures of uncertainty (each of which can be an LPS or an NPS) are equivalent if the preference order they place on random variables (according to their expected value) is the same. Note that, given an LPS $\vec{\mu}$, the expected value of a random variable X is $\sum_x x\vec{\mu}(X = x)$, where $\vec{\mu}(X = x)$ is a sequence of probability values and the multiplication and addition are pointwise. Thus, the expected value is a sequence; these sequences can be compared using the lexicographic order $<_L$ defined in Section 2.2. If ν is either an LPS or NPS, then let $E_\nu(X)$ denote the expected value of random variable X according to ν.

Definition 4.1: If each of ν_1 and ν_2 is either an NPS over $(W, \mathcal{F})$ or an LPS over $(W, \mathcal{F})$, then ν_1 is *equivalent to* ν_2, denoted $\nu_1 \approx \nu_2$, if, for all random variables X and Y measurable with respect to $\mathcal{F}$, $E_{\nu_1}(X) \leq E_{\nu_1}(Y)$ iff $E_{\nu_2}(X) \leq E_{\nu_2}(Y)$. (As usual, X is said to be measurable with respect to $\mathcal{F}$ if $\{w : X(w) = x\} \in \mathcal{F}$ for all x in the range of X.) ∎

This notion of equivalence satisfies analogues of the two key properties of the map $F_{S \to P}$ considered at the beginning of Section 3.

Proposition 4.2: *If $\nu \in \mathrm{NPS}(W, \mathcal{F})$, $\vec{\mu} \in \mathrm{LPS}(W, \mathcal{F})$, and $\nu \approx \vec{\mu}$, then $\nu(U) > 0$ iff $\vec{\mu}(U) > \vec{0}$. Moreover, if $\nu(U) > 0$, then $\mathrm{st}(\nu(V \mid U)) = \mu_j(V \mid U)$, where μ_j is the first probability measure in $\vec{\mu}$ such that $\mu_j(U) > 0$.*

The next result justifies restricting to finite LPS's if the state space is finite. Given an algebra $\mathcal{F}$, let *Basic*$(\mathcal{F})$ consist of the *basic sets* in $\mathcal{F}$, that is, the nonempty sets $\mathcal{F}$ that themselves contain no nonempty subsets in $\mathcal{F}$. Clearly the sets in *Basic*$(\mathcal{F})$ are disjoint, so that $|Basic(\mathcal{F})| \leq |W|$. If all sets are measurable, then *Basic*$(\mathcal{F})$ consists of the singleton subsets of W. If W is finite, it is easy to see that all sets in $\mathcal{F}$ are finite unions of the sets in *Basic*$(\mathcal{F})$.

Proposition 4.3: *If W is finite, then every LPS over $(W, \mathcal{F})$ is equivalent to an LPS of length at most $|Basic(\mathcal{F})|$.*

I can now define the isomorphism that relates NPS's and LPS's. Given $(W, \mathcal{F})$, let $LPS(W, \mathcal{F})/\!\approx$ be the equivalence classes of $\approx$-equivalent LPS's over $(W, \mathcal{F})$; similarly, let $NPS(W, \mathcal{F})/\!\approx$ be the equivalence classes of $\approx$-equivalent NPS's over $(W, \mathcal{F})$. Note that in $NPS(W, \mathcal{F})/\!\approx$, it is possible that different nonstandard probability measures could have different ranges. For this section, without loss of generality, I could also fix the range of all NPS's to be fixed nonstandard model $\mathbb{R}(\epsilon)$ discussed in Section 2.3. However, in the infinite case, it is not possible to restrict to a single nonstandard model, so I do not do so here either, for uniformity.

Now define the mapping $F_{L \to N}$ from $LPS(W, \mathcal{F})/\!\approx$ to $NPS(W, \mathcal{F})/\!\approx$ pretty much as suggested at the beginning of this subsection: If $[\vec{\mu}]$ is an equivalence class of LPS's, then choose a representative $\vec{\mu}' \in [\vec{\mu}]$ with finite length. Fix an infinitesimal ϵ. Suppose that $\vec{\mu}' = (\mu_0, \ldots, \mu_k)$. Let $F_{L \to N}([\vec{\mu}]) = [(1 - \epsilon - \cdots - \epsilon^k)\mu_0 + \epsilon\mu_1 + \cdots + \epsilon^k \mu_k]$.

Theorem 4.4: *If W is finite, then $F_{L \to N}$ is an isomorphism from $\mathrm{LPS}(W, \mathcal{F})/\!\approx$ to $\mathrm{NPS}(W, \mathcal{F})/\!\approx$ that preserves equivalence (that is, each NPS in $F_{L \to N}([\vec{\mu}])$ is equivalent to $\vec{\mu}$).*

BBD [1991a] also relate nonstandard probability measures and LPS's under the assumption that the state space is finite. However, the way they relate them is somewhat different in spirit from the notion of equivalence introduced here. They prove representation theorems essentially showing that a preference orders on lotteries can be represented by a standard utility function on lotteries and an LPS iff it can be represented by a standard utility function on lotteries and an NPS. Thus, they show that NPS's and LPS's are equiexpressive in terms of representing preference orders on lotteries.

The difference between BBD's result and Theorem 4.4 is essentially a matter of quantification. BBD's result can be viewed as showing that, given an LPS, for each utility function on lotteries, there is an NPS that generates the same preference order on lotteries for that particular utility function. In principle, the NPS might depend on the utility function. More precisely, for a fixed LPS $\vec{\mu}$, all that follows from their result is that for each utility function u, there is an NPS ν such that $(\vec{\mu}, u)$ and (ν, u) generate the same preference order on lotteries. Theorem 4.4 says that, given $\vec{\mu}$, there is an NPS ν such that $(\vec{\mu}, u)$ and (ν, u) generate the same preference on lotteries for *all* utility functions u.

4.2 The infinite case

An LPS over an infinite state space W may not be equivalent to any finite LPS. However, ideas analogous to those used to prove Proposition 4.3 can be used to provide a bound on the length of the minimal-length LPS's in an equivalence class.

Proposition 4.5: *Every LPS over $(W, \mathcal{F})$ is equivalent to an LPS over $(W, \mathcal{F})$ of length at most $|\mathcal{F}|$.*

Now, just as in the finite case, given an LPS $(\mu_\beta : \beta < \alpha)$ of length α, we want to show that it is equivalent to some NPS ν. Much like the finite case, the idea will be to take $\nu = \sum_{\beta<\alpha} \epsilon_\beta \mu_\beta$, where $st(\epsilon_{\beta'}/\epsilon_\beta) = 0$ if $\beta < \beta' < \alpha$. There are two issues that must be dealt with in order to get this to work. First, we must ensure that there is a non-Archimedean field where there are infinitesimals ϵ_β, $\beta < \alpha$, such that $st(\epsilon_{\beta'}/\epsilon_\beta) = 0$ if $\beta < \beta' < \alpha$. Note, for example, that this cannot be done in $I\!R(\epsilon)$ if $\alpha > \omega$. Another problem is making sense of the infinite sum. Fields are closed under finite sums; in general, infinite sums may not be defined.

I now construct a family of non-Archimedean fields where these problems are solved. Define a *nonstandard model of the integers* to be a model that contains the integers and satisfies every property of the integers expressible in first-order logic. It follows easily from the compactness theorem of first-order logic [Enderton 1972] that, given an ordinal α, there exists a nonstandard model of the integers that includes elements n_β, $\beta < \alpha$, such that $n_0 = 0$ and $n_\beta < n_{\beta'}$ if $\beta < \beta'$.[2]

Given a nonstandard model I^* of the integers, let $I\!R(I^*)$ be the non-Archimedean model defined as follows: $I\!R(I^*)$ consists of all polynomials of the form $\sum_{\beta<\alpha} r_\beta \epsilon^{n_\beta}$ for some ordinal

[2] The compactness theorem says that, given a collection for formulas, if each finite subset has a model, then so does the whole set. Consider a language with a function + and constant symbols for each integer, together with constants $\mathbf{n}_\beta$, $\beta < \alpha$. Consider the collection of first-order formulas in this language consisting of all the formulas true of the integers, together with the formulas $\mathbf{n}_0 = 0$ and $\mathbf{n}_\beta < \mathbf{n}_{\beta'}$, for all $\beta < \beta' < \alpha$. Clearly any finite subset of this set has a model—namely, the integers. Thus, by compactness, so does the full set. Clearly the model has the properties we want.

α, where $n_\beta \in I^*$ for $\beta < \alpha$, $n_\beta < n_{\beta'}$ if $\beta < \beta'$ (so that the set $\{n_\beta : \beta < \alpha\}$ is well founded), and r_β is a standard real for all $\beta < \alpha$. We can identify the standard real r with a polynomial of the form form $r\epsilon^0$. These polynomials can be added and multiplied using the standard rules for addition and multiplication of polynomials. It is easy to check that the result of adding or multiplying two polynomials is another polynomial in $I\!R(I^*)$. In particular, if p_1 and p_2 are two polynomials, N_1 is the set of coefficients of p_1, and N_2 is the set of coefficients of p_2, then the coefficients of $p_1 + p_2$ lie in $N_1 \cup N_2$, while the coefficients of $p_1 p_2$ lie in the set $N_3 = \{n_1 + n_2 : n \in N_1, n_2 \in N_2\}$. Both $N_1 \cup N_2$ and N_3 are easily seen to be well founded if N_1 and N_2 are. Moreover, for each expression $n_1 + n_2 \in N_3$, it follows from the well-foundedness of N_1 and N_2 that there are only finitely many pairs $(n, n') \in N_1 \times N_2$ such that $n + n' = n_1 + n_2$. Finally, each polynomial (other than 0) has an inverse that can be computed using standard "formal" division of polynomials; I leave the details to the reader. An element of $I\!R(I^*)$ is *positive* if its leading coefficient is positive. Define an order $\leq$ on $I\!R(I^*)$ by taking $a \leq b$ if $b - a$ is positive. With these definitions, $I\!R$ is a non-Archimedean field. Moreover, $st(\epsilon^{n_2}/\epsilon^{n_1}) = 0$ if $n_1 < n_2$.

Given $(W, \mathcal{F})$, let α be the minimal ordinal whose cardinality is greater than $|\mathcal{F}|$. Let $I^*_{(W,\mathcal{F})}$ be a nonstandard model of the integers such that there exist elements n_β in $I^*_{(W,\mathcal{F})}$ for all $\beta < \alpha$ such that $n_0 = 0$ and $n_\beta < n_{\beta'}$ if $\beta < \beta' < \alpha$. We can now define a map $F_{L \rightarrow N}$ from $LPS(W, \mathcal{F})/\!\approx$ to $NPS(W, \mathcal{F})/\!\approx$ as follows: Given an equivalence class $[\vec{\mu}] \in LPS(W, \mathcal{F})$, by Proposition 4.5, there exists $\vec{\mu} \in [\vec{\mu}]$ such that $\vec{\mu}$ has length $\alpha' \leq \alpha$. Let $\nu = \sum_{0 < \beta < \alpha'} \epsilon^{n_\beta} \mu_\beta$ and define $F_{L \rightarrow N}[\vec{\mu}] = [\nu]$. Arguments essentially identical to those of Lemma A.7 in the appendix can be used to show that $\nu \approx \vec{\mu}$. The following result is immediate.

Theorem 4.6: *$F_{L \rightarrow N}$ is an injection from* LPS$(W, \mathcal{F})/\!\approx$ *to* NPS$(W, \mathcal{F})/\!\approx$ *that preserves equivalence.*

What about the converse? Is it the case that for every NPS there is an equivalent LPS? As the following example shows, the answer is no.

Example 4.7: As in Example 3.2, let $W = I\!N$, the natural numbers, let $\mathcal{F}$ consist of the finite and cofinite subsets of $I\!N$, and let $\mathcal{F}' = \mathcal{F} - \{\emptyset\}$. Let ν^1 be an NPS with range $I\!R(\epsilon)$, where $\nu^1(U) = |U|\epsilon$ if U is finite and $\nu(U) = 1 - |\overline{U}|\epsilon$ if U is cofinite. This is clearly an NPS, and it corresponds to the cps μ^1 of Example 3.2, in the sense that $st(\nu(V \mid U)) = \mu^1(V \mid U)$ for all $V \in \mathcal{F}$, $U \in \mathcal{F}'$. Just as in Example 3.2, it can be shown that there is no LPS $\vec{\mu}$ such that $\nu^1 \approx \vec{\mu}$. ∎

4.3 Countably additive nonstandard probability measures

Do things get any better if we require countable additivity? To answer this question, we must first make precise what countable additivity means in the context of non-Archimedean fields. To understand the issue here, recall that for the standard real numbers, every bounded nondecreasing sequence has a unique least upper bound, which can be taken to be its limit. Given a countable sum each of whose terms is nonnegative, the partial sums form a nondecreasing sequence. If the partial sums are bounded (which they are if the terms in the sums represent the probabilities of a pairwise disjoint collection of sets), then the limit is well defined.

None of the above is true in the case of non-Archimedean fields. For a trivial counterexample, consider the sequence $\epsilon, 2\epsilon, 3\epsilon, \ldots$. Clearly this sequence is bounded (by any positive real number), but it does not have a least upper bound. For a more subtle example, consider the sequence $1/2, 3/4, 7/8, \ldots$ in the field $I\!R(\epsilon)$. Should its limit be 1? While this does not seem to be an unreasonable choice, note that 1 is not the least upper bound of the sequence. For example, $1-\epsilon$ is greater than every term in the sequence, and is less than 1. So are $1-3\epsilon$ and $1-\epsilon^2$. Indeed, this sequence has no least upper bound in $I\!R(\epsilon)$.

Despite these concerns, I define limits in $I\!R(I^*)$ pointwise. That is, a sequence $a_1, a_2, a_3, \ldots$ in $I\!R(I^*)$ converges to $b \in I\!R(I^*)$ if, for every $n \in I^*$, the coefficients of ϵ^n in $a_1, a_2, a_3, \ldots$ converge to the coefficient of ϵ^n in b. (Since the coefficients are standard reals, the notion of convergence for the coefficients is just the standard definition of convergence in the reals. Of course, if ϵ^n does not appear explicitly, its coefficient is taken to be 0.) As usual, $\sum_{i=1}^{\infty} a_i$ is taken to be b if the sequence of partial sums $\sum_{i=1}^{n} a_i$ converges to b. Note that, with this notion of convergence, $1/2, 3/4, 7/8, \ldots$ converges to 1 even though 1 is not the least upper bound of the sequence.[3] I discuss the consequences of this choice further in Section 7.

With this notion of countable sum, it makes perfect sense to consider countably-additive nonstandard probability measures. If $\mathcal{F}$ is a σ-algebra and $LPS^c(W, \mathcal{F})$ and $NPS^c(W, \mathcal{F})$ denote the countably additive LPS's and NPS's on $(W, \mathcal{F})$, respectively, then Proposition 4.6 can be applied with no change in proof to show the following.

Proposition 4.8: *$F_{L \to N}$ is an injection from* $\mathrm{LPS}^c(W, \mathcal{F})$ *to* $\mathrm{NPS}^c(W, \mathcal{F})$.

However, as the following example shows, even with the requirement of countable additivity, there are nonstandard probability measures that are not equivalent to any LPS.

Example 4.9: Let $W = \{w_1, w_2, w_3, \ldots\}$, and let $\mathcal{F} = 2^W$. Choose any nonstandard I^* and fix an infinitesimal ϵ in $I\!R(I^*)$. Define an NPS $(W, \mathcal{F}, \nu)$ with range $I\!R(I^*)$ by taking $\nu(w_j) = a_j + b_j\epsilon$, where $a_j = 1/2^j$, $b_{2j-1} = \epsilon/2^{j-1}$, and $b_{2j} = -\epsilon/2^{j-1}$, for $j = 1, 2, 3, \ldots$. Thus, the probabilities of $w_1, w_2, \ldots$ are characterized by the sequence $1/2 + \epsilon, 1/4 - \epsilon, 1/8 + \epsilon/2, 1/16 - \epsilon/2, 1/32 + \epsilon/4, \ldots$. For $U \subseteq W$, define $\nu(U) = \sum_{\{j : w_j \in U\}} a_j + \epsilon \sum_{\{j : w_j \in U\}} b_j$. It is easy to see that these sums are well-defined. As I show in the appendix (see Proposition A.9), there is no LPS $\vec{\mu}$ over $(W, \mathcal{F})$ such that $\nu \approx \vec{\mu}$. ∎

Roughly speaking, the reason that ν is not equivalent to any LPS in Example 4.9 is that the ratio between a_j and b_j in the definition of ν (i.e., the ratio "standard part" of $\nu(w_j)$ and the "infinitesimal part" of $\nu(w_j)$) grows unboundedly large. This can be generalized so as to give a condition on nonstandard probability measures that is necessary and sufficient to guarantee that they can be represented by an LPS. However, the condition is rather technical and I have not found an interesting interpretation of it, so I do not pursue it here.

[3]For those used to thinking of convergence in topological terms, what is going on here is that the topology corresponding to this notion of convergence is not Hausdorff.

5 Relating Popper Spaces to NPS's

Consider the map $F_{N\to P}$ from nonstandard probability spaces to Popper spaces such that $F_{N\to P}(W,\mathcal{F},\nu) = (W,\mathcal{F},\mathcal{F}',\mu)$, where $\mathcal{F}' = \{U : \nu(U) \neq 0\}$ and $\mu(V \mid U) = st(\nu(V \mid U))$ for $V \in \mathcal{F}$, $U \in \mathcal{F}'$. I leave it to the reader to check that $(W,\mathcal{F},\mathcal{F}',\mu)$ is indeed a Popper space. Define an equivalence relation $\simeq$ on $NPS(W,\mathcal{F})$ (and $NPS^c(W,\mathcal{F})$) by taking $\nu_1 \simeq \nu_2$ if $\{U : \nu_1(U) = 0\} = \{U : \nu_2(U) = 0\}$ and $st(\nu_1(V \mid U)) = st(\nu_2(V \mid U))$ for all V,U such that $\nu_1(U) \neq 0$. Let $NPS/\simeq$ (resp., $NPS^c/\simeq$) consist of the $\simeq$ equivalence classes in NPS (resp., NPS^c). Clearly $F_{N\to P}$ is well defined as a map from $NPS/\simeq$ to $Pop(W,\mathcal{F})$ and from $NPS^c/\simeq$ to $Pop^c(W,\mathcal{F})$. As the following result shows, $F_{N\to P}$ is actually a bijection.

Theorem 5.1: *$F_{N\to P}$ is a bijection from* $\mathrm{NPS}(W,\mathcal{F})/\simeq$ *to* $\mathrm{Pop}(W,\mathcal{F})/\simeq$ *and from* $\mathrm{NPS}^c(W,\mathcal{F})/\simeq$ *to* $\mathrm{Pop}^c(W,\mathcal{F})/\sim$.

McGee [1994] proves essentially the same result as Theorem 5.1 in the case that $\mathcal{F}$ is an algebra (and the measures involved are not necessarily countably additive). McGee [1994, p. 181] says that his result shows that "these two approaches amount to the same thing". However, this is far from clear. The $\simeq$ relation is rather coarse. In particular, it is coarser than $\approx$.

Proposition 5.2: *If $\nu_1 \approx \nu_2$ than $\nu_1 \simeq \nu_2$.*

The $\simeq$ relation identifies nonstandard measures that behave quite differently in decision contexts. This difference already arises in finite spaces, as the following example shows.

Example 5.3: Suppose $W = \{w_1, w_2\}$. Consider the nonstandard probability measure ν_1 such that $\nu_1(w_1) = 1/2 + \epsilon$ and $\nu_1(w_2) = 1/2 - \epsilon$. (This is equivalent to the LPS (μ_1, μ_2) where $\mu_1(w_1) = \mu_2(w_2) = 1/2$, $\mu_2(w_1) = 1$, and $\mu_2(w_2) = 0$.) Let ν_2 be the nonstandard probability measure such that $\nu_2(w_1) = \nu_2(w_2) = 1/2$. Clearly $\nu_1 \simeq \nu_2$. However, it is not the case that $\nu_1 \approx \nu_2$. Consider the two random variables $\chi_{\{w_1\}}$ and $\chi_{\{w_2\}}$. (I use the notation χ_U to denote the indicator function for U; that is, $\chi_U(w) = 1$ if $w \in U$ and $\chi_U(w) = 0$ otherwise.) According to ν_1, the expected value of $\chi_{\{w_1\}}$ is (very slightly) higher than that of $\chi_{\{w_2\}}$. According to ν_2, $\chi_{\{w_1\}}$ and $\chi_{\{w_2\}}$ have the same expected value. Thus, $\nu_1 \not\approx \nu_2$. Moreover, it is easy to see that there is no Popper measure μ on $\{w_1, w_2\}$ that can make the same distinctions with respect to $\chi_{\{w_1\}}$ and $\chi_{\{w_2\}}$ as ν_1, no matter how we define expected value with respect to a Popper measure. According to ν_1, although the expected value of $\chi_{\{w_1\}}$ is higher than that of $\chi_{\{w_2\}}$, the expected value of $\chi_{\{w_1\}}$ is less than that of $\alpha\chi_{\{w_2\}}$ for any (standard) real $\alpha > 1$. There is no Popper measure with this behavior. ∎

More generally, Theorem 3.1 shows that, in a precise sense, Popper spaces are equivalent to SLPS's, while Theorem 4.4 shows that LPS's are equivalent to NPS's. Thus, there is a gap in expressive power between Popper spaces and NPS's that essentially amounts to the gap between SLPS's and LPS's.

6 Independence

BBD [1991a] and Hammond discuss independence, but they consider only when a (standard or nonstandard) probability measure can be viewed as a *product measure* (that is, a product of other measures). Interestingly, their discussion does *not* consider independence directly for LPS's; indeed, it is far from clear what it would mean that an LPS can be written as a product measure. Rather than considering product measures, I start by considering more standard notions of independence: independence for events and then independence for random variables. The latter discussion will enable me to relate my definitions to those considered by BBD (and later Hammond).

Intuitively, event U is independent of V if learning U gives no information about V. Certainly if learning U gives no information about V, then if μ is an arbitrary probability measure, we would expect that $\mu(V \mid U) = \mu(V)$. Indeed, this is often taken as the definition of V being independent of U with respect to μ. If standard probability measures are used, conditioning is not defined if $\mu(U) = 0$. In this case, U is still considered independent of V. As is well known, if U is independent of V, then $\mu(U \cap V) = \mu(V) \times \mu(U)$ and V is independent of U, that is, $\mu(U \mid V) = \mu(U)$. Thus, independence of events with respect to a probability measure can be defined in any of three equivalent ways. Unfortunately, these definitions are not equivalent for other representations of uncertainty (see [Halpern 2003, Chapter 4] for a general discussion of this issue).

The situation is perhaps simplest for nonstandard probability measures.[4] In this case, the three notions coincide, for exactly the same reasons as they do for standard probability measures. However, independence is perhaps too strong a notion in some ways. In particular, nonstandard measures that are equivalent do not in general agree on independence, as the following example shows.

Example 6.1: Suppose that $W = \{w_1, w_2, w_3, w_4\}$. Let $\nu_i(w_1) = 1 - 2\epsilon + \epsilon_i$, $\nu_i(w_2) = \nu_i(w_3) = \epsilon - \epsilon_i$, and $\nu_i(w_4) = \epsilon_i$, for $i = 1, 2$, where $\epsilon_1 = \epsilon^2$ and $\epsilon_2 = \epsilon^3$. If $U = \{w_2, w_4\}$ and $V = \{w_3, w_4\}$, then $\nu_i(U) = \nu_i(V) = \epsilon$ and $\nu_i(U \cap V) = \epsilon_i$. It follows U and V are independent with respect to ν_1, but not with respect to ν_2. However, it is easy to check that $\nu_1 \approx \nu_2$. ∎

Example 6.1 shows that independence of events in the context of nonstandard measures is very sensitive to the choice of ϵ, even if this choice does not affect decision making at all. This suggests the following definition: U is *approximately independent* of V with respect to ν if $\nu(V \mid U) - \nu(V)$ is infinitesimal, that is, if $st(\nu(V \mid U)) = st(\nu(V))$. Note that U can be approximately independent of V without V being approximately independent of U. For example, consider the nonstandard probability measure ν_1 from Example 6.1. Let $V' = \{w_1, w_2\}$; as before, let $U = \{w_2, w_4\}$. It is easy to check that $st(\nu_1(V' \mid U)) = st(\nu_1(V')) = 1$, but $st(\nu_1(U \mid V')) = 1$, while $st(\nu_1(U)) = 0$. Thus, U is approximately independent of V' with respect to ν_1, but V' is not approximately independent of U. It is easy to check that, in general, U is approximately independent of V with respect to ν iff $st((\nu(V \cap U) - \nu(V) \times \nu(U))/\nu(U)) = 0$, while V is

[4] Although I talk about U being independent of V with respect to a nonstandard measure ν, technically I should talk about U being independent of V with respect to an NPS $(W, \mathcal{F}, \nu)$, for $U, V \in \mathcal{F}$. There seems to be no harm in being a little sloppy in the case of NPS's, although it will be a little more important to take the algebra into account in the case of Popper spaces.

approximately independent of U with respect to ν iff $st((\nu(V \cap U) - \nu(V) \times \nu(U))/\nu(V)) = 0$. Note for future reference that each of these requirements is stronger than just requiring that $st(\nu(V \cap U) - \nu(V) \times \nu(U)) = 0$. The latter requirement is automatically met, for example, if the probability of either U or V is infinitesimal.

The definition of (approximate) independence extends in a straightforward way to (approximate) conditional independence. U is conditionally independent of V given V' with respect to a (standard or nonstandard) probability measure ν if $\nu(V \mid U \cap V') = \nu(V \mid V')$ (where conditional independence is taken to hold by convention if $\nu(U \cap V') = 0$). Again, for probability, U is conditionally independent of V given V' iff V is conditionally independent of U given V' iff $\nu(V \cap U \mid V') = \nu(V \mid V') \times \nu(U \mid V')$. U is approximately conditionally independent of V given V' with respect to ν if $st(\nu(V \mid U \cap V')) = st(\nu(V \mid V'))$. If V' is taken to be W, the whole space, then (approximate) conditional independence reduces to (aproximate) independence.

The following proposition shows that, although independence is not preserved by equivalence, approximate indepence is.

Proposition 6.2: *If U is approximately conditionally independent of V given V' with respect to ν, and $\nu \approx \nu'$, then U is approximately conditionally independent of V given V' with respect to ν'.*

Proof: Suppose that $\nu \approx \nu'$. I claim that for all events U_1 and U_2 such that $\nu_1(U_2) \neq 0$, $st(\nu(U_1)/\nu(U_2)) = st(\nu'(U_1)/\nu'(U_2))$. For suppose that $st(\nu(U_1)/\nu(U_2)) = \alpha$. Then it easily follows that $E_\nu(\chi_{U_1}) < E_\nu(\alpha'\chi_{U_2})$ for all $\alpha' > \alpha$, and $E_\nu(\chi_{U_1}) > E_\nu(\alpha''\chi_{U_2})$ for all $\alpha'' < \alpha$. Thus, the same must be true for $E_{\nu'}$, and hence $st(\nu'(U_1)/\nu'(U_2)) = \alpha$. It thus follows that $st(\nu(V \mid U \cap V')) = st(\nu'(V \mid U \cap V'))$ and $st(\nu(V \mid V')) = st(\nu'(V \mid V'))$, from which the result is immediate. ∎

There is an obvious definition of independence for events for Popper spaces: U is independent of V given V' with respect to the Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$ if $U \cap V' \in \mathcal{F}'$ implies that $\mu(V \mid U \cap V') = \mu(V \mid V')$; if $V \cap V' \notin \mathcal{F}'$, then U is also taken to be independent of V given V'. If U is independent of V given V' and $V' \in \mathcal{F}'$, then $\mu(U \cap V \mid V') = \mu(U \mid V') \times \mu(V \mid V')$. However, the converse does not necessarily hold. Nor is it the case that if U is independent of V given V' then V is independent of U given V'. A counterexample can be obtained by taking the Popper space arising from the NPS in Example 6.1. Consider the Popper space $(W, 2^W, \mathcal{F}', \mu)$ corresponding to the NPS $(W, 2^W, \nu_1)$ via the isomorphism $F_{N \to P}$. It is easy to check that U is independent of V' but V' is not independent of U with respect to this Popper space, although $\mu(V' \cap U) = \mu(U \mid V') \times \mu(V')$ $(= 0)$. This observation is an instance of the following more general result, which is almost immediate from the definitions:

Proposition 6.3: *U is approximately independent of V given V' with respect to the NPS $(W, \mathcal{F}, \nu)$ iff U is independent of V given V' with respect to the Popper space $F_{N \to P}(W, \mathcal{F}, \nu)$.*

How should independence be defined in LPS's? Requiring that $\vec{\mu}(V \mid U) = \vec{\mu}(V)$ will not work since $\vec{\mu} \mid U$ and $\vec{\mu}$ are, in general, LPS's of different lengths. Nor is there any obvious way to define multiplication of two LPS's. It seems to me that the most natural way to define independence in LPS's is to essentially reduce the definition to that for Popper spaces. That is,

U is independent of V given V' with respect to the LPS $(W, \mathcal{F}, \vec{\mu})$ if the leftmost number in the sequence $\vec{\mu}(V \mid U \cap V')$ is the same as the leftmost number in $\vec{\mu}(V \mid V')$; as usual, independence is taken to hold trivially if $\vec{\mu}(U \cap V') = \vec{0}$. Again, the following result is almost immediate from the definitions.

Proposition 6.4: *U is independent of V given V' with respect to the LPS $\vec{\mu}$ iff U is approximately independent of V given V' with respect to each NPS in the equivalence class $F_{L \to N}([\vec{\mu}])$.*

Propositions 6.3 and 6.4 emphasize the naturalness of approximate independence in this context.

I now consider independence for random variables. Let $\mathcal{V}(X)$ denote the set of possible values (i.e., the range) of random variable X. For simplicity here, assume that the range of all random variables is finite. In the context of standard probability, random variable X is taken to be independent of Y if the event $X = x$ is independent of the event $Y = y$, for all $x \in \mathcal{V}(X)$ and $y \in \mathcal{V}(Y)$. It easily follows that if X is independent of Y, then $X \in U_1$ is independent of $Y \in V_1$ conditional on $Y \in V_2$ and $Y \in V_1$ is independent of $X \in U_1$ conditional on $X \in U_2$, for all $U_1, U_2 \subseteq \mathcal{V}(X)$ and $V_1, V_2 \subseteq \mathcal{V}(Y)$. The same arguments show that this is also true for for nonstandard probability measures. However, the argument breaks down for approximate independence.

Example 6.5: Suppose that $W = \{1,2,3\} \times \{1,2\}$. Let X and Y be the random variables that project onto the first and second components of a world, respectively, so that $X(i,j) = i$ and $Y(i,j) = j$. Let ν be the nonstandard probability measure on W given by the following table:

	1	2
1	$1 - 3\epsilon - 3\epsilon^2$	ϵ
2	ϵ	ϵ^2
3	ϵ	$2\epsilon^2$

It is easy to check that $X = i$ is approximately independent of $Y = j$ and that $Y = j$ is approximately independent of $X = i$ with respect to ν, for all $i \in \{1,2,3\}$, $j \in \{2,3\}$. However, $st(\nu(X = 2 \mid X \in \{2,3\} \cap Y = 2)) = 1/3$, while $st(\nu(X = 2 \mid X \in \{2,3\})) = 1/2$. ∎

In light of this example, I define X to be *approximately independent of Y with respect to ν* if $X \in U_1$ is approximately independent of $Y \in V_1$ conditional on $Y \in V_2$ for all $U_1 \subseteq \mathcal{V}(X)$ and $V_1, V_2 \subseteq \mathcal{V}(Y)$. I leave to the reader the obvious analogues of this definition for Popper spaces and LPS's.

Now I can compare the definitions given here to those discussed by BBD and Hammond. BBD define a (standard or nonstandard) probability measure ν on $W = W_1 \times \cdots \times W_n$ to be a product measure if there exist measures ν_i on W_i for $i = 1, \ldots, n$, such that such that $\nu((w_1, \ldots, w_n)) = \nu_1(w_1) \times \cdots \times \nu_n(w_n)$. If X_i is the random variable that projects on to the ith component, then it is easy to see that ν is a product measure iff $X_1, \ldots, X_n$ are independent. BBD then go on to define a decision-theoretic notion of stochastic independence on preference relations on acts over W. Under appropriate assumptions, it can be shown that a preference relation is stochastically independent iff it can be represented by some (real-valued) utility function u and a nonstandard probility measure ν such that $X_1, \ldots, X_n$ are approximately independent with respect to ν [Battigalli and Veronesi 1996]. BBD also consider a weak notion of product measure that requires only that there exist measures $\nu_1, \ldots, \nu_n$ such that $st((\nu(w_1, \ldots, w_n)) = st(\nu_1(w_1) \times \cdots \nu(w_n))$. As we have already observed, this notion of independence is rather weak. Indeed, an example in BBD shows that it misses out on some interesting decision-theoretic behavior.

Hammond mainly focuses on Popper spaces, follows BBD's lead in considering when a Popper space can be, in a sense, viewed as a product measure. He defines a notion of conditional independece of a Popper space defined on $W = W_1 \times \cdots \times W_n$ which is similar in spirit to the notion of independence of random variables in Popper spaces as defined here. In fact, it is straightforward to show that the Popper space $(W_1 \times \cdots \times W_n, \mathcal{F}, \mathcal{F}', \mu)$ is conditionally independent in Hammond's sense iff the random variables $X_1, \ldots, X_n$ are independent with respect to the Popper space, in the sense defined here.

7 Discussion

As the preceding discussion shows, there is a sense in which NPS's are more general than both Popper spaces and LPS's. LPS's are more expressive than Popper measures in finite spaces and in infinite spaces where we assume countable additivity (in the sense discussed at the end of Section 5), but without assuming countable additivity, they are incomparable, as Examples 3.2 and 3.3 show.

Although NPS's are the most expressive of the three approaches I have considered, they have some disadvantages. In particular, working with a nonstandard probability measure requires defining and working with a non-Archimedean field. LPS's have the advantage of using just standard probability measures. Moreover, their lexicographic structure may give useful insights. It seems to be worth considering the extent to which LPS's can be generalized so as to increase their expressive power. I am currently exploring LPS's ordered by an arbitrary (not necessarily well-founded) index set. It seems that such LPS's may be useful in characterizing iterated deletion of weakly dominated strategies. (This is done by Brandenburger and Keisler [2000] using finite LPS's; it seems that results are more cleanly stated using infinite LPS's ordered by the integers.) I hope to report on this in future work.

I conclude with a brief discussion of a few other issues raised by this paper.

- Belief: The connections between LPS's, NPS's, and CPS's are relevant to the notion of belief. Brandenburger and Keisler [2000] defined a notion of belief using LPS's and provided an elegant decision-theoretic justification of it. According to their definition, an agent *believes U in LPS $\vec{\mu}$* if there is some $j \leq m$ such that $\mu_i(U) = 1$ for all $i \leq j$ and $\mu_i(U) = 0$ for $i > j$. Independently, van Fraassen [1995] defined a notion of belief using Popper spaces that can be shown to be essentially equivalent to the definition given by Brandenburger and Keisler. That there should be equivalent notions of belief in the context of LPS's and Popper spaces is perhaps not that surprising, in light of the close connection between them. The results of this paper suggest that it may also be worth considering notions of belief defined in NPS's.

- Nonstandard utility: In this paper I have assumed that, while probabilities may be lexicographically ordered or nonstandard, utilities are standard real numbers. There is a tradition in decision theory going back to Hausner [1954] and continued recently in a sequence of papers by Fishburn and Lavalle (see [Fishburn and Lavalle 1998] and the references therein) of considering nonstandard or lexicographically-ordered utilities. I have not considered the relationship between these ideas and the ones considered here, but there may be some fruitful connections.

- Countable additivity for NPS's: Countable additivity for standard probability measures is essentially a continuity condition. The fact that $\sum_{i=1}^{\infty} a_i$ may not be the least upper bound of the partial sums $\sum_{i=1}^{n} a_i$ in an NPS leads to a certain lack of continuity in decision-making. For example, let $W = \{w_1, w_2, \ldots\}$. Consider a nonstandard probability measure ν such that $\nu(w_1) = 1/3 - \epsilon$, $\nu(w_2) = 1/3 + \epsilon$, and $\nu(w_{k+2}) = 1/(3 \times 2^k)$, for $k = 1, 2, \ldots$. Let $U_n = \{w_3, \ldots, w_n\}$ and let $U_\infty = \{w_3, w_4, \ldots\}$. Clearly $\nu(U_n) \rightarrow \nu(U_\infty) = 1/3$. However, $\nu(U_n) < \nu(w_1)$ for all n. Thus, $E_\nu(\chi_{\{w_1\}}) > E_\nu(\chi_{U_n})$ for all $n \geq 3$ although $E_\nu(\chi_{\{w_1\}}) < E_\nu(\chi_{U_\infty})$.

  Not surprisingly, the same situations can be modeled with LPS's. Consider the LPS (μ_1, μ_2), where $\mu_1 = st(\nu_1)$, $\mu(w_1) = 0$, $\mu_2(w_2) = 2/3$, and $\mu_2(w_{k+2}) = 1/(3 \times 2^k)$ for $k = 1, 2, \ldots$. It is easy to see that again $E_{\vec{\mu}}(\chi_{\{w_1\}}) > E_{\vec{\mu}}(\chi_{U_n})$ for all $n \geq 3$ although $E_{\vec{\mu}}(\chi_{\{w_1\}}) < E_\nu(\chi_{U_\infty})$. (A similar example can be obtained using SLPS's, by replacing each world w_i by a pair of worlds w_i', w_i'', where w_i' is in the support of μ_1 and w_i'' of μ_2.)

  An analogous continuity problem arises even in finite domains. Let $W = \{w_1, w_2, w_3\}$ and consider a sequence of probability measures ν_n such that $\nu_n(w_1) = 1/3 - 1/n$, $\nu_n(w_2) = 1/3 - \epsilon$ and $\nu(w_3) = 1/3 + 1/n + \epsilon$. Clearly $\nu_n \rightarrow \nu$, where $\nu(w_1) = 1/3$, $\nu(w_2) = 1/3 - \epsilon$, and $\nu(w_3) = 1/3 + \epsilon$. However, $\nu_n(\chi_{\{w_1\}}) < \nu_n(\chi_{\{w_2\}})$ for all n, while $\nu(\chi_{\{w_1\}}) > \nu(\chi_{\{w_2\}})$. Again, the same situation can be modeled using LPS's (and even SLPS's).

  Is this lack of continuity a problem? I am not sure, but I believe it deserves further thought.

Acknowledgments: I'd like to thank Adam Brandenburger and Peter Hammond for a number of very enlightening discussions, Bob Stalnaker for pointing out Example 3.2, Brain Skyrms for pointing me to Hammond's work, Bas van Fraassen for pointing me to Spohn's work, and Larry Blume, Eddie Dekel, and Horacio Arlo-Costa for a number of useful comments on an earlier draft of this paper.

A Appendix: Proofs

In this section, I prove all the results claimed in the main part of the paper. For the convenience of the reader, I repeat the statements of the results.

Theorem 3.1: *If W is finite, the map $F_{S \rightarrow P}$ is an isomorphism from* SLPS$(W, \mathcal{F})$ *to* Pop$(W, \mathcal{F})$.

Proof: The only difficulty comes in showing that $F_{S \rightarrow P}$ is onto. Given a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$, we must show that there is an SLPS $(W, \mathcal{F}, \vec{\mu})$ that is mapped to $(W, \mathcal{F}, \mathcal{F}', \mu)$ by

$F_{S\to P}$. Inductively define a sequence of pairwise disjoint sets in $\mathcal{F}'$ as follows. Let $U_0 = \emptyset$. Let U_1 be the smallest set $U \in \mathcal{F}$ such that $\mu(U) = 1$. Since W is finite, there is such a smallest set (it is simply the intersection of all sets U such that $\mu(U) = 1$). Since $\mu(U_1 \mid W) > 0$, it follows that $U_1 \in \mathcal{F}'$. If $\overline{U}_1 \in \mathcal{F}'$ and $\overline{U}_1 \neq \emptyset$, let U_2 be the smallest set in $\mathcal{F}$ such that $\mu(U_2 \mid \overline{U}_1) = 1$. Note that $U_2 \in \mathcal{F}'$. Continuing in this way, it is clear that there exists a sequence of sets $U_0, U_1, \ldots, U_k$ such that (1) $U_i \in \mathcal{F}'$ for $i = 1, \ldots, k$, (2) for $i < k$, $\overline{U_0 \cup \ldots \cup U_i} \in \mathcal{F}'$ and U_{i+1} is the smallest subset of $\mathcal{F}$ such that $\mu(U_{i+1} \mid \overline{U_0 \cup \ldots \cup U_i}) = 1$ and (3) either $\overline{U_1 \cup \ldots \cup U_k} \notin \mathcal{F}'$ or $\overline{U_1 \cup \ldots \cup U_k} = \emptyset$. Note that condition (2) guarantees U_{i+1} is a subset of $\overline{U_0 \cup \ldots \cup U_i}$), so the U_i's are pairwise disjoint. Define the LPS $\vec{\mu} = (\mu_1, \ldots, \mu_k)$ by taking $\mu_i(V) = \mu(V \mid U_i)$. Clearly the support of μ_i is U_i, so this is an LCPS (and hence an SLPS).

To see that $(W, \mathcal{F}, \mathcal{F}', \mu)$ is the result of applying $F_{S\to P}$ to $(W, \mathcal{F}, \vec{\mu})$, suppose that instead $(W, \mathcal{F}, \mathcal{F}'', \mu')$ is the result. I first show that $\mathcal{F}' = \mathcal{F}''$. Suppose that $V \in \mathcal{F}''$. Then $\mu_i(V) > 0$ for some i. Thus, $\mu(V \mid U_i) > 0$. Since $U_i \in \mathcal{F}'$, it follows that $V \in \mathcal{F}'$. Thus, $\mathcal{F}'' \subseteq \mathcal{F}'$.

To show that $\mathcal{F}' \subseteq \mathcal{F}''$, first note that, by construction, $\mu(U_j \mid \overline{U_1 \cup \ldots \cup U_{j-1}}) = 1$. It easily follows that if $V \subseteq \overline{U_1 \cup \ldots \cup U_{j-1}}$ then

$$\mu(V \mid \overline{U_1 \cup \ldots \cup U_{j-1}}) = \mu(V \cap U_j \mid \overline{U_1 \cup \ldots \cup U_{j-1}}).$$

Thus, by CP3,

$$\mu(V \mid \overline{U_1 \cup \ldots \cup U_{j-1}}) = \mu(V \cap U_j \mid \overline{U_1 \cup \ldots \cup U_{j-1}}) = \mu(V \mid U_j) \times \mu(U_j \mid \overline{U_1 \cup \ldots \cup U_{j-1}}),$$

so

$$\mu(V \mid U_j) = \mu(V \mid \overline{U_1 \cup \ldots \cup U_{j-1}}). \tag{1}$$

Now suppose that $V \in \mathcal{F}'$. Clearly $V \cap (U_1 \cup \ldots \cup U_k) \neq \emptyset$, for otherwise $V \subseteq \overline{U_1 \cup \ldots \cup U_k}$, contradicting the fact that $\overline{U_0 \cup \ldots \cup U_k} \notin \mathcal{F}'$. Let j_V be the smallest index j such that $V \cap U_j \neq \emptyset$. Since $U_0 = \emptyset$, $j_V > 0$. I claim that $\mu(V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) \neq 0$. For if $\mu(V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) = 0$, then $\mu(U_{j_V} - V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) = 1$, contradicting the definition of U_{j_V} as the smallest set U' such that $\mu(U' \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) = 1$. Moreover, since $V \subseteq \overline{U_1 \cup \ldots U_{j_V-1}}$, it follows from (1) that $\mu(V \mid U_{j_V}) = \mu(V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) > 0$. Thus, $\mu_{j_V}(V) > 0$, so $V \in \mathcal{F}''$.

This argument can be extended to show that $\mu(V' \mid V) = \mu'(V' \mid V)$ for all $V' \in \mathcal{F}$. Since $V \cap U_i = \emptyset$ for $i < j_V$, it follows that $\mu'(V' \mid V) = \mu_{j_V}(V' \mid V)$. By CP3, $\mu(V' \mid V) \times \mu(V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}}) = \mu(V' \cap V \mid \overline{U_0 \cup \ldots \cup U_{j_V-1}})$. By (1) and the fact that $\mu(V \mid U_{j_V}) > 0$, it follows that $\mu(V' \mid V) = \mu(V' \cap V \mid U_{j_V}) / \mu(V \mid U_{j_V})$, i.e., that $\mu(V' \mid V) = \mu_{j_V}(V' \mid V)$. ∎

Although Theorem 3.4 was proved by Spohn [1986], I include a proof here as well, to make the paper self-contained.

Theorem 3.4: *For all W, the map $F_{S\to P}$ is an isomorphism from $\mathrm{SLPS}^c(W, \mathcal{F})$ to $\mathrm{Pop}^c(W, \mathcal{F})$.*

Proof: Again, the difficulty comes in showing that $F_{S\to P}$ is onto. Given a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$, I again show that there is an LCPS $(W, \mathcal{F}, \vec{\mu})$ that is mapped to $(W, \mathcal{F}, \mathcal{F}', \mu)$ by $F_{S\to P}$. However, now a completely different proof is required. The earlier inductive construction of the sequence $U_0, \ldots, U_k$ no longer works. The problem already arises in the construction of

U_1. There may no longer be a smallest set U_1 such that $\mu(U_1) = 1$. Consider, for example, the interval $[0, 1]$ with Borel measure. There is clearly no smallest subset U of $[0, 1]$ such that $\mu(U) = 1$.

As a first step to getting around this, put an order $\leq$ on sets in $\mathcal{F}'$ by defining $U \leq V$ if $\mu(U \,|\, U \cup V) > 0$. (Essentially, the same order is considered by van Fraassen [1976].)

Lemma A.1: *$\leq$ is transitive.*

Proof: By definition, if $U \leq V$ and $V \leq V'$, then $\mu(U \,|\, U \cup V) > 0$ and $\mu(V \,|\, V \cup V') > 0$. To see that $\mu(U \,|\, U \cup V') > 0$, note that $\mu(U \,|\, U \cup V \cup V') + \mu(V \,|\, U \cup V \cup V') + \mu(V' \,|\, U \cup V \cup V') = 1$, so at least one of $\mu(U \,|\, U \cup V \cup V')$, $\mu(V \,|\, U \cup V \cup V')$, or $\mu(V' \,|\, U \cup V \cup V')$ is positive. I consider each of the cases separately.

Case 1: Suppose that $\mu(U \,|\, U \cup V \cup V') > 0$. By CP3,

$$\mu(U \,|\, U \cup V \cup V') = \mu(U \,|\, U \cup V') \times \mu(U \cup V' \,|\, U \cup V \cup V').$$

Thus, $\mu(U \,|\, U \cup V') > 0$, as desired.

Case 2: Suppose that $\mu(V \,|\, U \cup V \cup V') > 0$. By assumption, $\mu(U \,|\, U \cup V) > 0$; since $\mu(V \,|\, U \cup V \cup V') > 0$, it follows that $\mu(U \cup V \,|\, U \cup V \cup V') > 0$. Thus, by CP3,

$$\mu(U \,|\, U \cup V \cup V') = \mu(U \,|\, U \cup V) \times \mu(U \cup V \,|\, U \cup V \cup V') > 0.$$

Thus, case 2 can be reduced to case 1.

Case 3: Suppose that $\mu(V' \,|\, U \cup V \cup V') > 0$. By assumption, $\mu(V \,|\, V \cup V') > 0$; since $\mu(V' \,|\, U \cup V \cup V') > 0$, it follows that $\mu(V \cup V' \,|\, U \cup V \cup V') > 0$. Thus, by CP3,

$$\mu(V \,|\, U \cup V \cup V') = \mu(V \,|\, V \cup V') \times \mu(V \cup V' \,|\, U \cup V \cup V') > 0.$$

Thus, case 3 can be reduced to case 2.

This completes the proof, showing that $\leq$ is transitive. ∎

Define $U \sim V$ if $U \leq V$ and $V \leq U$.

Lemma A.2: *$\sim$ is an equivalence relation on $\mathcal{F}'$.*

Proof: It is immediate from the definition that $\sim$ is reflexive and symmetric; transitivity follows from the transitivity of $\leq$. ∎

Rényi [1956] and van Fraassen [1976] also considered the $\sim$ relation in their papers, and the argument that $\leq$ is transitive is similar in spirit to Rényi's argument that $\sim$ is transitive. However, the rest of this proof diverges from those of Rényi and van Fraassen.

Let $[U]$ denote the $\sim$-equivalence class of U, and let $\mathcal{F}'/{\sim} = \{[U] : U \in \mathcal{F}'\}$.

Lemma A.3: *Each equivalence class $[V] \in \mathcal{F}'/\!\sim$ is closed under countable unions.*

Proof: Suppose that $V_1, V_2, \ldots \in [V]$. I must show that $\cup_{i=1}^{\infty} V_i \in [V]$. Clearly $V_j \leq \cup_{i=1}^{\infty} V_i$ for all j. Suppose, by way of contradiction, that $\cup_{i=1}^{\infty} V_i \not\leq V_j$ for some j. Since $\leq$ is transitive, it follows that $V_j < \cup_{i=1}^{\infty} V_i$ for all j. Thus, $\mu(V_j \mid \cup_{i=1}^{\infty} V_i) = 0$ for all j. But then, by countable additivity,

$$1 = \mu(\cup_{i=1}^{\infty} V_i) \mid \cup_{i=1}^{\infty} V_i) \leq \sum_{j=1}^{\infty} \mu(V_j \mid \cup_{i=1}^{\infty} V_i) = 0,$$

a contradiction. Thus, $[V]$ is closed under countable unions. ∎

Fix an element $V_0 \in [V]$.

Lemma A.4: $\inf\{\mu(V_0 \mid V_0 \cup V') : V' \in [V]\} > 0$.

Proof: Suppose that $\inf\{\mu(V_0 \mid V_0 \cup V') : V' \in [V]\} = 0$. Then there exist sets $V_1, V_2, \ldots$ such that $\mu(V_0 \mid V_0 \cup V_n) < 1/n$. Since $[V]$ is closed under countable unions, $\cup_{i=1}^{n} V_i \in [V]$. Since $V_0 \sim \cup_{i=1}^{n} V_i$, it follows that $\mu(V_0 \mid \cup_{i=0}^{\infty} V_i) > 0$. But, by CP3,

$$\mu(V_0 \mid \cup_{i=0}^{\infty} V_i) = \mu(V_0 \mid V_0 \cup V_n) \times \mu(V_0 \cup V_n \mid \cup_{i=0}^{\infty} V_i) \leq \mu(V_0 \mid V_0 \cup V_n) \leq 1/n.$$

Since this is true for all $n > 0$, it follows that $\mu(V_0 \mid \cup_{i=0}^{\infty} V_i) = 0$, a contradiction. ∎

The next lemma shows that each equivalence class in $\mathcal{F}'/\!\sim$ has a “maximal element”.

Lemma A.5: *In each equivalence class $[V]$, there is an element $V^* \in [V]$ such that $\mu(V^* \mid V' \cup V^*) = 1$ for all $V' \in [V]$.*

Proof: Again, fix an element $V_0 \in [V]$. By Lemma A.4, there exists some $\alpha_V > 0$ such that $\inf\{\mu(V_0 \mid V_0 \cup V') : V' \in [V]\} = \alpha_V$. Thus, there exist sets $V_1, V_2, V_3, \ldots \in [V]$ such that $\mu(V_0 \mid V_0 \cup V_n) < \alpha + 1/n$. By Lemma A.3, $V^* = \cup_{i=0}^{\infty} V_i \in [V]$. By CP3,

$$\mu(V_0 \mid V^*) = \mu(V_0 \mid V_0 \cup V_n) \times \mu(V_0 \cup V_n \mid V^*) \leq \mu(V_0 \mid V_0 \cup V_n) < \alpha_V + 1/n.$$

Thus, $\mu(V_0 \mid V^*) \leq \alpha_V$. By choice of α_V, it follows that $\mu(V_0 \mid V^*) = \alpha_V$.

Suppose that $\mu(V^* \mid V' \cup V^*) < 1$ for some $V' \in [V]$. But then, by CP3,

$$\mu(V_0 \mid V' \cup V^*) = \mu(V_0 \mid V^*) \times \mu(V^* \mid V' \cup V^*) < \alpha_V,$$

contradicting the choice of α_V. Thus, $\mu(V^* \mid V' \cup V^*) = 1$ for all $V' \in [V]$. ∎

Define a total order on these equivalence relations by taking $[U] \leq [V]$ if $U' \leq V'$ for some $U' \in [U]$ and $V' \in [V]$. It is easy to check (using the transitivity of $\leq$) that if $U' \leq V'$ for some $U' \in [U]$ and some $V' \in [V]$, then $U'' \leq V''$ for all $U'' \in [U]$ and all $V'' \in [V]$.

Lemma A.6: *$\leq$ is a well-founded relation on $\mathcal{F}'/\!\sim$.*

Proof: Note that if $[U] < [V]$, then $\mu(V \mid U \cup V) = 0$. It now follows from countable additivity that $<$ is a well-founded order on these equivalence classes. For suppose that there exists an infinite decreasing sequence $[U_0] > [U_1] > [U_2] > \ldots$. Since $\mathcal{F}$ is a σ-algebra, $\cup_{i=0}^{\infty} U_i \in \mathcal{F}$; since $\mathcal{F}'$ is closed under supersets, $\cup_{i=0}^{\infty} U_i \in \mathcal{F}'$. By CP3,

$$\mu(U_j \mid \cup_{i=0}^{\infty} U_i) = \mu(U_j \mid U_j \cup U_{j+1}) \times \mu(U_j \cup U_{j+1} \mid \cup_{i=0}^{\infty} U_i) = 0.$$

Let $V_0 = U_0$ and, for $j > 0$, let $V_j = U_j - (\cup_{i=0}^{j-1} U_j)$. Clearly the V_j's are pairwise disjoint, $\cup_i U_i = \cup_i V_i$, and $\mu(V_j \mid \cup_{i=0}^{\infty} U_i) \leq \mu(U_j \mid \cup_{i=0}^{\infty} U_i) = 0$. It now follows that using countable additivity that

$$1 = \mu(\cup_{i=0}^{\infty} U_i \mid \cup_{i=0}^{\infty} U) = \sum_{i=0}^{\infty} \mu(V_i \mid \cup_{i=0}^{\infty} U_i) = 0.$$

This is as contradiction, so the equivalence classes are well-founded. ∎

Because $\leq$ is well-founded, there is an order-preserving isomorphism O from $\mathcal{F}'/\!\sim$ to an initial segment of the ordinals (i.e., $[U] \leq [V]$ iff $O([U]) \leq O([V])$). Thus, the equivalence classes can be enumerated using all the ordinals less than some ordinal α. By Lemma A.5, there are sets U_β, $\beta < \alpha$, in $\mathcal{F}'$ such that if $O([U]) = \beta$, then $U_\beta \in [U]$ and $\mu(U_\beta \mid U \cup U_\beta) = 1$ for all $U' \in [U]$. Define an LPS $\vec{\mu} = (\mu_0, \mu_1, \ldots)$ of length α by taking $\mu_\beta(V) = \mu(V \mid U_\beta)$. The choice of the U_β's guarantees that this is actually an SLPS.

It remains to show that $(W, \mathcal{F}, \mathcal{F}', \mu)$ is the result of applying $F_{C \rightarrow P}$ to $(W, \mathcal{F}, \vec{\mu})$. Suppose that instead $(W, \mathcal{F}, \mathcal{F}'', \mu')$ is the result. The argument that $\mathcal{F}'' \subseteq \mathcal{F}'$ is identical to that in the finite case: If $V \in \mathcal{F}''$, then $\mu_\beta(V) > 0$ for some β. Thus, $\mu(V \mid U_\beta) > 0$. Since $U_\beta \in \mathcal{F}'$, it follows that $V \in \mathcal{F}'$. Thus, $\mathcal{F}'' \subseteq \mathcal{F}'$.

Now suppose that $V \in \mathcal{F}'$. Thus, $V \sim V_\beta$ for some $\beta < \alpha$. It follows that $\mu(V \mid V_\beta) > 0$, so $V \in \mathcal{F}''$.

Finally, to show that $\mu(U \mid V) = \mu'(U \mid V)$, suppose that β is such that $V \sim V_\beta$. It follows that $\mu(V \mid V_{\beta'}) = 0$ for $\beta' < \beta$ and $\mu(V \mid V_\beta) > 0$. Thus, by definition, $\mu'(U \mid V) = \mu_\beta(U \mid V)$. Without loss of generality, assume that $U \subseteq V$ (otherwise replace U by $U \cap V$). Thus, by CP3,

$$\mu(U \mid V) \times \mu(V \mid V \cup V_\beta) = \mu(U \mid V \cup V_\beta). \tag{2}$$

Suppose $V' \subseteq V$. Clearly

$$\mu(V' \mid V \cup V_\beta) = \mu(V' \cap V_\beta \mid V \cup V_\beta) + \mu(V' \cap \overline{V_\beta} \mid V \cup V_\beta).$$

Now by CP3 and the fact that $\mu(V_\beta \mid V \cup V_\beta) = 1$,

$$\mu(V' \cap V_\beta \mid V \cup V_\beta) = \mu(V' \mid V_\beta) \times \mu(V_\beta \mid V \cup V_\beta) = \mu(V' \mid V_\beta)$$

and

$$\mu(V' \cap \overline{V_\beta} \mid V \cup V_\beta) \leq \mu(\overline{V_\beta} \mid V \cup V_\beta) = 0.$$

Thus, $\mu(V' \mid V \cup V_\beta) = \mu(V' \mid V_\beta)$. Applying this observation to both U and V shows that $\mu(V \mid V \cup V_\beta) = \mu(V \mid V_\beta)$ and $\mu(U \mid V \cup V_\beta) = \mu(U \mid V_\beta)$. Plugging this into (2), it follows that

$$\mu(U \mid V) = \mu(U \mid V_\beta)/\mu(V \mid V_\beta) = \mu_\beta(U)/\mu_\beta(V) = \mu_\beta(U \mid V) = \mu'(U \mid V).$$

This completes the proof of the theorem. ∎

Proposition 4.2: *If $\nu \approx \vec{\mu}$, then $\nu(U) > 0$ iff $\vec{\mu}(U) > \vec{0}$. Moreover, if $\nu(U) > 0$, then* $\mathrm{st}(\nu(V \mid U)) = \mu_j(V \mid U)$, *where μ_j is the first probability measure in $\vec{\mu}$ such that $\mu_j(U) > 0$.*

Proof: Recall that for $U \subseteq W$, χ_U is the indicator function for U; that is, $\chi_U(w) = 1$ if $w \in U$ and $\chi_U(w) = 0$ otherwise. Notice that $E_\nu(\chi_U) > E_\nu(\chi_\emptyset)$ iff $\nu(U) > 0$ and $E_{\vec{\mu}}(\chi_U) > E_{\vec{\mu}}(\chi_\emptyset)$ iff $\vec{\mu}(U) > \vec{0}$. Since $\nu \approx \vec{\mu}$, it follows that $\nu(U) > 0$ iff $\vec{\mu}(U) > \vec{0}$. If $\nu(U) > 0$, note that $E_\nu(\chi_{U\cap V} - \alpha\chi_U) > E_\nu(\chi_\emptyset)$ iff $\alpha < st(\nu(V \mid U))$. Similarly, $E_{\vec{\mu}}(\chi_{U\cap V} - \alpha\chi_U) > E_{\vec{\mu}}(\chi_\emptyset)$ iff $\alpha < \mu_j(U)$, where j is the least index such that $\mu_j(U) > 0$. It follows that $st(\nu(V \mid U)) = \mu_j(V \mid U)$. ∎

Proposition 4.3: *If W is finite, then every LPS over $(W, \mathcal{F})$ is equivalent to an LPS of length at most $|Basic(\mathcal{F})|$.*

Proof: Suppose that W is finite and $Basic(\mathcal{F}) = \{U_1, \ldots, U_k\}$. Given an LPS $\vec{\mu}$, define a finite subsequence $\vec{\mu}' = (\mu_{m_0}, \ldots, \mu_{m_h})$ of $\vec{\mu}$ as follows. Let $\mu_{k_0} = \mu_0$. Suppose that $\mu_{k_0}, \ldots, \mu_{k_j}$ have been defined. If all probability measures in $\vec{\mu}$ with index greater that k_j are linear combinations of the probability measures with index $\mu_{k_0}, \ldots, \mu_{k_j}$, then take $\vec{\mu}' = (\mu_{k_0}, \ldots, \mu_{k_j})$. Otherwise, let $\mu_{k_{j+1}}$ be the probability measure in $\vec{\mu}$ with least index that is not a linear combination of $\mu_{k_0}, \ldots, \mu_{k_j}$. Since a probability measure over $(W, \mathcal{F})$ is determined by its value on the sets in $Basic(\mathcal{F})$, a probability measure over $(W, \mathcal{F})$ can be identified with a vector in $I\!R^{|Basic(\mathcal{F})|}$: the vector defining the probabilities of the elements in $Basic(\mathcal{F})$. There can be at most $|Basic(\mathcal{F})|$ linearly independent such vectors, thus $\vec{\mu}'$ has length at most $|Basic(\mathcal{F})|$.

It remains to show that $\vec{\mu}'$ is equivalent to $\vec{\mu}$. Given random variables X and Y, suppose that $E_{\vec{\mu}}(X) < E_{\vec{\mu}}(Y)$. Then there is some minimal index β such that $E_{\mu_\gamma}(X) = E_{\mu_\gamma}(Y)$ for all $\gamma < \beta$ and $E_{\mu_\beta}(X) < E_{\mu_\beta}(Y)$. It follows that μ_β cannot be a linear combination of μ_γ for $\gamma < \beta$. Thus, μ_β is one of the probability measures in $\vec{\mu}'$. Moreover, the expected value of X and Y agree for all probability measures in $\vec{\mu}'$ with lower index (since they do in $\vec{\mu}$). Thus, $E_{\vec{\mu}'}(X) < E_{\vec{\mu}'}(X)$.

The argument in the other direction is similar in spirit and left to the reader. ∎

Theorem 4.4: *If W is finite, then $F_{L\to N}$ is an isomorphism from* $\mathrm{LPS}(W, \mathcal{F})/\!\approx$ *to* $\mathrm{NPS}(W, \mathcal{F})/\!\approx$ *that preserves equivalence (that is, each NPS in $F_{L\to N}([\vec{\mu}])$ is equivalent to $\vec{\mu}$).*

Proof: I first prove a general characterization of when an NPS is equivalent to an LPS.

Lemma A.7: *Suppose that $\epsilon_0, \ldots, \epsilon_k$ are such that* $\mathrm{st}(\epsilon_{i+1}/\epsilon_i) = 0$ *for $i = 1, \ldots, k-1$ and $\sum_{i=0}^k \epsilon_i = 1$. Then $(\mu_0, \ldots, \mu_k) \approx \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$.*

Proof: Let $\vec{\mu} = (\mu_0, \ldots, \mu_k)$ and let $\nu = \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$. Suppose that $E_{\vec{\mu}}(X) < E_{\vec{\mu}}(Y)$. Thus, there exists some $j \leq k$ such that $E_{\mu_j}(X) < E_{\mu_j}(Y)$ and $E_{\mu_{j'}}(X) = E_{\mu_{j'}}(Y)$ for all $j' < j$. Since $E_\nu(X) = \sum_{i=0}^k \epsilon_i E_{\mu_i}(X)$ and $E_\nu(Y) = \sum_{i=0}^k \epsilon_i E_{\mu_i}(Y)$, to show that $E_\nu(X) < E_\nu(Y)$, it suffices to show that $\epsilon_j(E_{\mu_j}(X) - E_{\mu_j}(Y)) > \sum_{i=j+1}^k \epsilon_i(E_{\mu_i}(X) - E_{\mu_i}(Y))$. Since $\epsilon_{j'+1} \leq \epsilon_{j'}$ for $j' \geq j$, it follows that $\sum_{i=j+1}^k \epsilon_i(E_{\mu_i}(X) - E_{\mu_i}(Y)) \leq \epsilon_{j+1} \sum_{i=j+1}^k |E_{\mu_i}(X) - E_{\mu_i}(Y)|$. Thus, it suffices to show that $\epsilon_{j+1}/\epsilon_j < (E_{\mu_j}(X) - E_{\mu_j}(Y))/\sum_{i=j+1}^k |E_{\mu_i}(X) - E_{\mu_i}(Y)|$. Since the right-hand side of the inequation is a positive real and $st(\epsilon_{j+1}/\epsilon_j) = 0$, the result follows.

The argument in the opposite direction is similar. Suppose that $E_\nu(X) < E_\nu(Y)$. Again, since $E_\mu(X) = \sum_{i=0}^k \epsilon_i E_{\mu_i}(X)$ and $E_\mu(Y) = \sum_{i=0}^k \epsilon_i E_{\mu_i}(Y)$, it must be the case that if j is the least index such that $E_{\mu_j}(X) \neq E_{\mu_j}(Y)$, then $E_{\mu_j}(X) < E_{\mu_j}(Y)$. Thus, $E_{\vec{\mu}}(X) < E_{\vec{\mu}}(Y)$. ∎

It follows from Lemma A.7 that

$$\vec{\mu}' = (\mu_0, \ldots, \mu_k) \approx (1 - \epsilon - \cdots - \epsilon^k)\mu_0 + \epsilon\mu_1 + \cdots + \epsilon^k \mu_k.$$

It remains to show that, given an NPS $(W, \mathcal{F}, \nu)$, there is an equivalence class $[\vec{\mu}]$ such that $F_{L\to N}([\vec{\mu}]) = [\nu]$. My goal is to find (standard) probability measures $\mu_0, \ldots, \mu_k$ and $\epsilon_0, \ldots, \epsilon_k$ such that $st(\epsilon_{i+1}/\epsilon_i) = 0$ and $\nu = \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$. If this can be done then, by Lemma A.7, $\nu \approx (\mu_0, \ldots, \mu_k)$, and we are done.

Suppose that $\mathcal{F}$ has a basis $U_1, \ldots, U_k$ and that ν has range $I\!R^*$. Note that a probability measure ν' on $\mathcal{F}$ can be identified with a vector $(a_1, \ldots, a_k)$ over $I\!R^*$, where $\nu'(U_i) = a_i$, so that $a_1 + \cdots + a_k = 1$. In the rest of this proof, I frequently identify ν with such a vector.

Lemma A.8: *There exist* $k' \leq k$, $\epsilon_0, \ldots, \epsilon_{k'}$ *where* $\epsilon_0 = 1$, $\mathrm{st}(\epsilon_{i+1}/\epsilon_i) = 0$ *for* $i = 1, \ldots, k'-1$, *and standard real-valued vectors* $\vec{b}_j$, $j = 0, \ldots, k'$, *in* $I\!R^k$ *such that*

$$\nu = \sum_{j=0}^{k'} \epsilon_j \vec{b}_j.$$

Proof: I show by induction on $m \leq k$ that there exist $\epsilon_0, \ldots, \epsilon_m$ and $m' \leq m$ such that $\epsilon_j = 0$ for $j' > m'$, $st(\epsilon_{i+1}/\epsilon_i) = 0$ for $i = 1, \ldots, m'-1$, and standard vectors $\vec{b}_j$ $j = 0, \ldots, m-1$ and a possibly nonstandard vector $\vec{b}'_m = (b'_{m1}, \ldots, b'_{mk})$ such that (a) $\nu = \sum_{j=0}^{m-1} \epsilon_j \vec{b}_j + \epsilon_m \vec{b}'_m$, (b) $|b'_{mi}| \leq 1$, and (c) at least m of $b'_{m1}, \ldots, b'_{mk}$ are standard.

For the base case (where $m = 0$), just take $\vec{b}'_0 = \nu$ and $\epsilon_0 = 1$. For the inductive step, suppose that $0 < m < k$. If $\vec{b}'_m$ is standard, then take $\vec{b}_m = \vec{b}'_m$, $\vec{b}_{m+1} = \vec{0}$, and $\epsilon_{m+1} = 0$. Otherwise, let $\vec{b}_m = st(\vec{b}'_m)$ and let $\vec{b}''_{m+1} = \vec{b}'_m - \vec{b}_m$. Let $\epsilon' = \max\{|b''_{(m+1)i}| : i = 1, \ldots, k\}$. Since not all components of $\vec{b}'_m$ are standard, $\epsilon' > 0$. Note that, by construction, $st(\epsilon'/b_{mi}) = 0$ if $b_{mi} \neq 0$, for $i = 1, \ldots, k$. Let $\vec{b}'_{m+1} = \vec{b}''_{m+1}/\epsilon'$ and let $\epsilon_{m+1} = \epsilon'\epsilon_m$. By construction, $|b'_{(m+1)i}| \leq 1$ and at least one component of $\vec{b}'_{m+1}$ is either 1 or -1. Moreover, if b'_{mi} is standard, then $b''_{(m+1)i} = b'_{(m+1)i} = 0$. Thus, $\vec{b}'_{m+1}$ has at least one more standard component that $\vec{b}'_m$. Since clearly $\nu = \sum_{j=0}^m \epsilon_j \vec{b}_j + \epsilon_{m+1} \vec{b}'_{m+1}$, this completes the inductive step. The lemma follows immediately. ∎

Returning to the proof of Theorem 4.4, I next prove by induction on m that for all $m \leq k'$ (where $k' \leq k$ is as in Lemma A.8), there exist standard probability measures $\mu_0, \ldots, \mu_m$, (standard) vectors $\vec{b}_{m+1}, \ldots, \vec{b}_{k'} \in I\!R^k$, and $\epsilon_1, \ldots, \epsilon_{k'}$ such that $\nu = \sum_{j=0}^m \epsilon_j \mu_j + \sum_{j=m+1}^{k'} \epsilon_j \vec{b}_j$.

The base case is immediate from Lemma A.8: taking $\vec{b}_j$, $j = 1, \ldots, k'$ as in Lemma A.8, $\vec{b}_0$ is in fact a probability measure since $\vec{b}_0 = st(\nu)$. Suppose that the result holds for m. Consider $\vec{b}_{m+1}$. If $b_{(m+1)i} < 0$ for some j then, since $\nu(U_i) \geq 0$, there must exist $j' \in \{1, \ldots, m\}$ such that $\mu_{j'}(U_i) > 0$. Thus, there exists some $N > 0$ such that $N(\mu_{j'}(U_i)) + b_{(m+1)i} > 0$. Since there are only finitely many basic elements and every element in the vector μ_j is nonnegative, for $j = 0, \ldots, m$, there must exist some N' such that $\vec{b}'_{m+1} = N'(\mu_0 + \cdots + \mu_m) + \vec{b}_{m+1} \geq 0$. Let $c = \sum_{i=1}^{k} b'_{(m+1)i}$, and let $\mu_{m+1} = \vec{b}'_{m+1}/c$. Clearly, $\nu = (\epsilon_0 - N'\epsilon_{m+1})\mu_0 + \cdots (\epsilon_m - N'\epsilon_{m+1})\mu_m + c\epsilon_{m+1}\mu_{m+1} + \sum_{j=m+2}^{k'} \vec{b}_j$. This completes the proof of the inductive step.

The theorem now immediately follows. ∎

Proposition 4.5: *Every LPS over $(W, \mathcal{F})$ is equivalent to an LPS over $(W, \mathcal{F})$ of length at most $|\mathcal{F}|$.*

Proof: The argument is essentially the same as that for Proposition 4.3, using the observation that a probability measure over $(W, \mathcal{F})$ can be identified with an element of $\mathbb{R}^{|\mathcal{F}|}$; the vector defining the probabilities of the elements in $\mathcal{F}$. I leave details to the reader. ∎

Proposition A.9: *For the NPS $(W, \mathcal{F}, \nu)$ constructed in Example 4.9, there is no LPS $\vec{\mu}$ over $(W, \mathcal{F})$ such that $\nu \approx \vec{\mu}$.*

Proof: I start with a straightforward lemma.

Lemma A.10: *Given an LPS $\vec{\mu}$, there is an LPS $\vec{\mu}'$ such that $\vec{\mu} \approx \vec{\mu}'$ and all the probability measures in $\vec{\mu}'$ are distinct.*

Proof: Define $\vec{\mu}'$ to be the subsequence consisting of all the distinct probability measures in $\vec{\mu}$. That is, suppose that $\vec{\mu} = (\mu_0, \mu_1, \ldots)$. Then $\vec{\mu} = (\mu_{k_0}, \mu_{k_1}, \ldots)$, where $k_0 = 0$, and, if k_α has been defined for all $\alpha < \beta$ and there exists an index γ such that $\mu_{k_\alpha} \neq \mu_\gamma$ for all $\alpha \leq \beta$, then k_β is the least index δ such that $\mu_{k_\alpha} \neq \mu_\delta$ for all $\alpha < \beta$. If there is no index γ such that $\mu_\gamma \notin \{\mu_{k_\alpha} : \alpha < \beta\}$, then $\vec{\mu}' = (\mu_{k_\alpha} : \alpha < \beta)$. I leave it to the reader to check that $\vec{\mu} \approx \vec{\mu}'$. ∎

Returning to the proof of Proposition A.9, suppose by way of contradiction that $\nu \approx \vec{\mu}$. Without loss of generality, by Lemma A.10, assume that all the probability measures in $\vec{\mu}$ are distinct. Clearly $E_\nu(\chi_W) < E_\nu(\alpha\chi_{\{w_1\}})$ if $\alpha \geq 2$ and $E_\nu(\chi_W) > E_\nu(\alpha\chi_{\{w_1\}})$ if $\alpha < 2$. Since $\nu \approx \vec{\mu}$, it must be the case that $E_{\vec{\mu}}(\chi_W) < E_{\vec{\mu}}(\alpha\chi_{\{w_1\}})$ if $\alpha \geq 2$ and $E_{\vec{\mu}}(\chi_W) > E_{\vec{\mu}}(\alpha\chi_{\{w_1\}})$ if $\alpha < 2$. Since $E_{\vec{\mu}}(\chi_W) = (1, 1, \ldots)$, it follows that if $\vec{\mu} = (\mu_0, \mu_1, \ldots)$, it must be the case that $\mu_0(w_1) = 1/2$ and

$$\mu_1(w_1) \geq 1/2. \tag{3}$$

Similar arguments (comparing χ_W to $\chi_{\{w_j\}}$) can be used to show that $\mu_0(w_j) = 1/2^j$ and $\mu_1(w_{2j-1}) \geq 1/2^j$ for $j = 1, 2, \ldots$. Next, observe that $E_\nu(\chi_{\{w_1\}} - 2^{2k-1}\chi_{\{w_{2k}\}}) = (2^k + 1)\epsilon$. Thus,

$$E_\nu(\chi_{\{w_1\}} - 2^{2k-1}\chi_{\{w_{2k}\}}) = E_\nu((2^k + 1)(\chi_{\{w_1\}} - (\chi_W/2))).$$

It follows that the same relationship must hold if ν is replaced by $\vec{\mu}$. That is,

$$\mu_1(w_1) - 2^{2k-1}\mu_1(w_{2k}) = (2^k+1)(\mu_1(w_1) - (1/2)).$$

Rearranging terms, this gives

$$2^k\mu_1(w_1) + 2^{2k-1}\mu(w_{2k}) = 2^{k-1} + 1/2,$$

or

$$\mu_1(w_1) + 2^{k-1}\mu(w_{2k}) = 1/2 + 1/2^{k+1}. \tag{4}$$

Thus, $\mu_1(w_1) \leq 1/2 + 1/2^{k+1}$ for all $k \geq 1$. Putting this together with (3), it follows that $\mu_1(w_1) = 1/2$. Plugging this into (4) gives $\mu_1(w_{2k}) = 1/2^{2k}$. It now follows that $\mu_1 = \mu_0$, contradicting the choice of $\vec{\mu}$. ∎

Theorem 5.1: *$F_{N\to P}$ is a surjection from* NPS$(W,\mathcal{F})$ *to* Pop$(W,\mathcal{F})$ *and from* NPS$^c(W,\mathcal{F})$ *to* Pop$^c(W,\mathcal{F})$*.*

Proof: The result in the case that W is countably additive is immediate from Theorems 3.4, Proposition 4.5, and Proposition 4.2. Thus, it remains to prove the result in the case that W is infinite and $\mathcal{F}$ is an algebra (but not necessarily a σ-algebra). McGee [1994] proves a result essentially equivalent to Theorem 5.1 in this case. My proof follows the lines of his. I provide the details here mainly for completeness.

The proof relies on the following ultrafilter construction of non-Archimedean fields. Given a set S, a *filter* $\mathcal{G}$ on S is a nonempty set of subsets of $\mathcal{F}$ that is closed under supersets (so that if $U \in \mathcal{G}$ and $U \subseteq U'$, then $U' \in \mathcal{G}$), is closed under finite intersections (so that if $U_1, U_2 \in \mathcal{G}$, then $U_1 \cap U_2 \in \mathcal{G}$), and does not contain $\emptyset$. An *ultrafilter* is a maximal filter, that is, a filter that is not a strict subset of any other filter. It is not hard to show that if $\mathcal{U}$ is an ultrafilter on S, then for all $U \subseteq S$, either $U \in \mathcal{U}$ or $\overline{U} \in \mathcal{U}$ [Bell and Slomson 1974].

Suppose F is either $I\!R$ or a non-Archimedean field, J is an arbitrary set, and $\mathcal{U}$ is an ultrafilter on J. Define an equivalence relation $\sim_{\mathcal{U}}$ on F^J by taking $(a_j : j \in J) \sim_{\mathcal{U}} (b_j : j \in J)$ if $\{j : a_j = b_j\} \in \mathcal{U}$. Similarly, define a total order $\preceq_{\mathcal{U}}$ by taking $(a_j : j \in J) \preceq_{\mathcal{U}} (b_j : j \in J)$ if $\{j : a_j \leq b_j\} \in \mathcal{U}$. (The fact that $\leq_{\mathcal{U}}$ is total uses the fact that for all $U \subseteq J$, either $U \in \mathcal{U}$ or $\overline{U} \in \mathcal{U}$. Note that the pointwise ordering on F^J is not total.) Let $F^J/{\sim_{\mathcal{U}}}$ consist of these equivalence classes. Note that F can be viewed as a subset of $F^J/{\sim_{\mathcal{U}}}$ by identifying $a \in F$ with the sequence of all a's.

Define addition and multiplication on F^J pointwise, so that, for example, $(a_j : j \in J) + (b_j : j \in J) = (a_j + b_j : j \in J)$. It is easy to check that if $(a_j : j \in J) \sim_{\mathcal{U}} (a'_j : j \in J)$, then $(a_j : j \in J) + (b_j : j \in J) \sim_{\mathcal{U}} (a'_j : j \in J) + (b_j : j \in J)$, and similarly for multiplication. Thus, the definitions of $+$ and $\times$ can be extended in the obvious way to $F^J/{\sim_{\mathcal{U}}}$. With these definitions, it is easy to check that $F^J/{\sim_{\mathcal{U}}}$ is a field that contains F.

Now given a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$ and a finite subset $\mathcal{A} = \{U_1, \ldots, U_k\} \subseteq \mathcal{F}$, let $\mathcal{F}_{\mathcal{A}}$ be the (finite) algebra generated by $\mathcal{A}$ (that is, the smallest set containing $\{U_1, \ldots, U_k, W\}$ that is closed under unions and complement). Let $\mathcal{F}'_{\mathcal{A}} = \mathcal{F}_{\mathcal{A}} \cap \mathcal{F}'$. It follows from Theorem 3.1 that there is a finite SLPS $\vec{\mu}_{\mathcal{A}}$ over $(W, \mathcal{F}_{\mathcal{A}})$ that is mapped to $(W, \mathcal{F}_{\mathcal{A}}, \mathcal{F}'_{\mathcal{A}'}, \mu_{\mathcal{A}})$ by $F_{S\to P}$.

(Although Theorem 3.1 is stated for finite state spaces W, the proof only uses the fact that the algebra is finite, so it applies without change here.) It now follows from Theorem 4.4 that, for each $\mathcal{A}$, there is a nonstandard probability space $(W, \mathcal{F}_\mathcal{A}, \nu_\mathcal{A})$ with range $I\!R(\epsilon)$ that is equivalent to $\vec{\mu}_\mathcal{A}$. By Proposition 4.2, it follows that for $U \in \mathcal{F}'_\mathcal{A}$ iff $\nu_\mathcal{A}(U) = 0$. Moreover, $st(\nu_\mathcal{A}(V \mid U)) = \mu_\mathcal{A}(V \mid U)$ for $U \in \mathcal{F}'_\mathcal{A}$ and $V \in \mathcal{F}_\mathcal{A}$.

Let J consist of all finite subsets of $\mathcal{F}$. For a subset $\mathcal{A}$ of $\mathcal{F}$, let $G_\mathcal{A}$ be the subset of 2^J consisting of all sets in J containing $\mathcal{A}$. Let $\mathcal{G} = \{G \subseteq J : G \supseteq G_\mathcal{A}$ for some $\mathcal{A} \subseteq \mathcal{F}\}$. It is easy to check that $\mathcal{G}$ is a filter on J. It is a standard result that every filter can be extended to an ultrafilter [Bell and Slomson 1974]. Let $\mathcal{U}$ be an ultrafilter containing $\mathcal{G}$. By the construction above, $\mathcal{R}(\epsilon)/{\sim_\mathcal{U}}$ is a non-Archimedean field.

Define ν on $(W, \mathcal{F})$ by taking $\nu(U) = (\nu_\mathcal{A}(U) : \mathcal{A} \in J)$, where $\nu_\mathcal{A}(U)$ is taken to be 0 if $U \notin \mathcal{F}_\mathcal{A}$. To see that ν is indeed a nonstandard probability measure with the required properties, note that clearly $\nu(W) = 1$ (where 1 is identified with the sequence of all 1's). Moreover, to see that $\nu(U) + \nu(V) = \nu(U \cup V)$, let $\mathcal{A}_{U,V}$ be the smallest subalgebra containing U and V. Note that if $\mathcal{A} \supset \mathcal{A}_{U,V}$, then $\nu_\mathcal{A}(U) + \nu_\mathcal{A}(V) = \nu_{\mathcal{A}'}(U \cup V)$. Since the set of algebras containing $\mathcal{A}_{U,V}$ is an element of the ultrafilter, the result follows. Similar arguments show that $\nu(U) = 0$ iff $U \in \mathcal{F}'$ and that $st(\nu(V \mid U)) = \mu(V \mid U)$ if $U \in \mathcal{F}'$ and $V \in \mathcal{F}$. Clearly $F_{N \to P}(\nu) = \mu$. ∎

Proposition 5.2: *If $\nu_1 \approx \nu_2$ than $\nu_1 \simeq \nu_2$.*

Proof: Suppose that $\nu_1 \approx \nu_2$. To show that $\nu_1 \simeq \nu_2$, first suppose that $\nu_1(U) \neq 0$ for some $U \subseteq W$. Then $E_{\nu_1}(\chi_\emptyset) < E_{\nu_1}(\chi_U)$. Since $\nu_1 \approx \nu_2$, it must be the case that $E_{\nu_2}(\chi_\emptyset) < E_{\nu_2}(\chi_U)$. Thus, $\nu_2(U) \neq 0$. A symmetric argument shows that if $\nu_2(U) \neq 0$ then $\nu_1(U) \neq 0$. Next, suppose that $\nu_1(U) \neq 0$ and $\nu_1(V \mid U) = \alpha$. Thus, $E_{\nu_1}(\alpha\chi_U) = E_{\nu_1}(\chi_{U \cap V})$. Since $\nu_1 \approx \nu_2$, it follows that $E_{\nu_2}(\alpha\chi_U) = E_{\nu_2}(\chi_{U \cap V})$, and so $\nu_2(V \mid U) = \alpha$. Thus, $st(\nu_1(V \mid U)) = st(\nu_2(V \mid U))$. Hence, $\nu_1 \simeq \nu_2$, as desired. ∎

References

Adams, E. (1966). Probability and the logic of conditionals. In J. Hintikka and P. Suppes (Eds.), *Aspects of Inductive Logic*, pp. 265–316. Amsterdam: North-Holland.

Battigalli, P. and P. Veronesi (1996). A note on stochastic independence without Savage-null events. *Journal of Economic Theory 70*(1), 235–248.

Bell, J. L. and A. B. Slomson (1974). *Models and Ultraproducts: An Introduction (third revised printing)*. Amsterdam/New York: North-Holland/American Elsevier.

Blume, L., A. Brandenburger, and E. Dekel (1991a). Lexicographic probabilities and choice under uncertainty. *Econometrica 59*(1), 61–79.

Blume, L., A. Brandenburger, and E. Dekel (1991b). Lexicographic probabilities and equilibrium refinements. *Econometrica 59*(1), 81–98.

Brandenburger, A. and J. Keisler (2000). Epistemic conditions for iterated admissibility. Unpublished manuscript.

Császár, A. (1955). Sur la structure des espace de probabilité conditionelle. *Acta Mathematica Academiae Scientiarum Hungaricae 6*, 337–361.

Davis, M. (1977). *Applied Nonstandard Analysis*. New York: Wiley.

de Finetti, B. (1936). Les probabilités nulles. *Bulletins des Science Mathématiques (première partie) 60*, 275–288.

Enderton, H. B. (1972). *A Mathematical Introduction to Logic*. New York: Academic Press.

Fishburn, P. C. and I. H. Lavalle (1998). Subjective expected lexicographic utility with infinite state sets. *Journal of Mathematical Economics 30*(3), 323–346.

Halpern, J. Y. (2003). *Reasoning About Uncertainty*. Cambridge, Mass.: MIT Press.

Hammond, P. J. (1994). Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick Suppes: Scientific Philosopher; Volume 1*. Dordrecht, Netherlands: Kluwer.

Hausner, M. (1954). Multidimensional utilities. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision Processes*, pp. 167–180. New York: Wiley.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Kreps, D. M. and R. B. Wilson (1982). Sequential equilibria. *Econometrica 50*, 863–894.

Lehmann, D. and M. Magidor (1992). What does a conditional knowledge base entail? *Artificial Intelligence 55*, 1–60.

McGee, V. (1994). Learning the impossible. In E. Eells and B. Skyrms (Eds.), *Probability and Conditionals*. Cambridge, U.K.: Cambridge University Press.

Myerson, R. (1986). Multistage games with communication. *Econometrica 54*, 323–358.

Popper, K. R. (1968). *The Logic of Scientific Discovery* (2nd ed.). London: Hutchison. The first version of this book appeared as *Logik der Forschung*, 1934.

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 in *Selected Papers of Alfred Rényi, I: 1948–1956*, pp. 554–557. Budapest: Akadémia Kiadó, 1976.

Rényi, A. (1964). Sur les espaces simples de probabilités conditionelles. *Annales de l'Institut Henri Poincaré, Nouvelle série, Section B 1*, 3–21. Reprinted as paper 237 in *Selected Papers of Alfred Rényi, III: 1962–1970*, pp. 284–302. Budapest: Akadémia Kiadó, 1976.

Robinson, A. (1973). Function theory on some nonarchimedean fields. *American Mathematical Monthly: Papers on the Foundations of Mathematics 80*, S87–S109.

Selten, R. (1965). Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit. *Zeitschrift für Gesamte Staatswissenschaft 121*, 301–324 and 667–689.

Spohn, W. (1986). The representation of Popper measures. *Topoi 5*, 69–74.

van Fraassen, B. C. (1976). Representation of conditional probabilities. *Journal of Philosophical Logic 5*, 417–430.

van Fraassen, B. C. (1995). Fine-grained opinion, probability, and the logic of full belief. *Journal of Philosophical Logic 24*, 349–377.

Wilson, N. (1995). An order of magnitude calculus. In *Proc. Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pp. 548–555.