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

November 11, 2009

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.

JEL classification numbers: C70; D80; D81;

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

*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. A preliminary version appeared in the Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge, 2001 [Halpern 2001]. This version includes detailed proofs and more discussion and more examples; in addition, the material in Section 6 (on independence) is new.

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 consideration of events of measure 0 may seem to be of little practical interest, it turns out to play a critical role in game theory, particularly in the analysis of strategic reasoning in extensive-form games and in the analysis of weak dominance in normal-form games (see, for example, [Battigalli 1996; Battigalli and Siniscalchi 2002; Blume, Brandenburger, and Dekel 1991a; Blume, Brandenburger, and Dekel 1991b; Brandenburger, Friedenberg, and Keisler 2008; Fudenberg and Tirole 1991; Hammond 1994; Hammond 1999; Kohlberg and Reny 1997; Kreps and Wilson 1982; Myerson 1986; Selten 1965; Selten 1975]). It also arises in the analysis of conditional statements by philosophers (see [Adams 1966; McGee 1994]), and in dealing with nonmonotonicity in Artificial Intelligence (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 involves *conditional probability spaces* (CPS's). The idea, which goes back to Popper [1934, 1968] and de Finetti [1936], is to take as primitive not probability, but conditional probability. If μ is a conditional probability measure on a space W, 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 $\mu(U \mid W) = 0$. A second approach, which goes back to at least Robinson [1973] and has been explored in the economics literature [Hammond 1994; Hammond 1999], 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 a third 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* (LPS's) 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 3.4). 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. One way to understand LPS's is in terms of NPS's. 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 $\mu_1(U) > 0$.

What is the precise relationship between these approaches? The relationship between LPS's and CPS's 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* (LCPS's) if the state space is finite and it is possible to condition on any nonempty set.[1] 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 where the events on which conditioning is allowed satisfy certain regularity conditions. As I show, this result depends critically on a number of assumptions. In particular, it does not work without the assumption of countable additivity, it requires that we extend LCPS's appropriately to the infinite case, and it is sensitive to the choice of conditioning events. For example, if we consider CPS's where the conditioning events can be viewed as information sets, and so are are not closed under supersets (this is essentially the case considered by Battigalli and Sinischalchi [2002]), then the result no longer holds.

Turning to the relationship between LPS's and NPS's, I show that if the state space is finite, then LPS's are in a sense equivalent to NPS's. More precisely, say that two measures of uncertainty ν_1 and ν_2 (each of which can be either an LPS or an NPS) are equivalent, denoted $\nu_1 \approx \nu_2$, if they cannot be distinguished by (real-valued) random variables; that is, for all random variables X and Y, $E_{\nu_1}(X) \leq E_{\nu_1}(Y)$ iff $E_{\nu_2}(X) \leq E_{\nu_2}(Y)$ (where $E_\nu(X)$ denotes the expected value of X with respect to ν). To the extent that we are interested in these representations of uncertainty for decision making, then we should not try to distinguish two representations that are equivalent. I show that, in finite spaces, there is a straightforward bijection between $\approx$-equivalence classes of LPS's and NPS's. 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.)

These results give some useful insight into independence of random variables. There have been a number of alternative notions of independence considered in the literature of extended probability spaces (i.e., approaches that deal with the problem of conditioning on sets of measure 0): BBD considered three; Kohlberg and Reny [1997] considered two others. It turns out that these notions are perhaps best understood in the context of NPS's; I describe and compare them here.

Many of the new results in this paper involve infinite spaces. Given that most games studied by game theorists are finite, it is fair to ask whether these results have any significance for game theory. I believe they do. Even if the underlying game is finite, the set of types is infinite. Epistemic characterizations of solution concepts often make use of *complete* type spaces, which include every possible type of every player, where a type determines an (extended) probability over the strategies and types of the other players; this must be an infinite space. For example, Battigalli and Siniscalchi [2002] use a complete type space where the uncertainty is represented by cps's to give an epistemic characteri-

[1]Despite this isomorphism; it is not clear that conditional probability spaces are *equivalent* to LPS's. It depends on exactly what we mean by equivalence. The same comment applies below where the word "equivalent" is used. See Section 7 for further discussion. I thank Geir Asheim for bringing this point to my attention.

zation of extensive-form rationalizability and backward induction, while Brandenburger, Friedenberg, and Keisler [2008] use a complete type space where the uncertainty is represented by LPS's to get a characterization of weak dominance in normal-form games. As the results of this paper show, the set of types depends to some extent on the notion of extended probability used. Similarly, a number of characterizations of solution concepts depend on independence (see, for example, [Battigalli 1996; Kohlberg and Reny 1997; Battigalli and Siniscalchi 1999]). Again, the results of this paper show that these notions can be somewhat sensitive to exactly how uncertainty is represented, even with a finite state space. While I do not present any new game-theoretic results here, I believe that the characterizations I have provided may be useful both in terms of defending particular choices of representation used and suggesting new solution concepts.

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 second argument comes from a set $\mathcal{F}'$ of conditioning events, that is, that is, events on which conditioning is allowed. One natural choice is to take $\mathcal{F}'$ to be $\mathcal{F} - \emptyset$. But it may be reasonable to consider other restrictions on $\mathcal{F}'$. For example, Battigalli and Sinischalchi [2002] take $\mathcal{F}'$ to consist of the information sets in a game, since they are interested only in agents who update their beliefs conditional on getting some information. The question is what constraints, if any, should be placed on $\mathcal{F}'$. For most of this paper, I focus on *Popper spaces* (named after Karl Popper), defined next, where the set $\mathcal{F}'$ satisfies four arguably reasonable requirements, but I occasionally consider other requirements (see Section 3.3).

Definition 2.1: A *conditional probability space (cps) over* $(W, \mathcal{F})$ is a tuple $(W, \mathcal{F}, \mathcal{F}', \mu)$ such that $\mathcal{F}$ is an algebra over W, $\mathcal{F}'$ is a set of subsets of W (not necessarily an algebra over W) that does not contain $\emptyset$, 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 $V \subseteq X \subseteq U$, $U, X \in \mathcal{F}'$, $V \in \mathcal{F}$.

Note that it follows from CP1 and CP2 that $\mu(\cdot \mid U)$ is a probability measure on $(W, \mathcal{F})$ (and, in particular, that $\mu(\emptyset \mid U) = 0$) for each $U \in \mathcal{F}'$. A *Popper space over* $(W, \mathcal{F})$ is a conditional probability space $(W, \mathcal{F}, \mathcal{F}', \mu)$ that satisfies three additional conditions: (a) $\mathcal{F}' \subseteq \mathcal{F}$, (b) $\mathcal{F}'$ is closed under supersets in $\mathcal{F}$, in that if $V \in \mathcal{F}'$, $V \subseteq V'$, and $V' \in \mathcal{F}$, then $V' \in \mathcal{F}'$, and (c) 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, I use a superscript c to denote the restriction to countably additive Popper spaces, so $Pop^c(W, \mathcal{F})$ denotes 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 last 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. Note that, since this regularity condition depends on the Popper measure, it may well be the case that $(W, \mathcal{F}, \mathcal{F}', \mu)$ and $(W, \mathcal{F}, \mathcal{F}', \nu)$ are both cps's over $(W, \mathcal{F})$, but only the former is a Popper space over $(W, \mathcal{F})$.

Popper [1934, 1968] and de Finetti [1936] were the first to formally consider conditional probability as the basic notion, although as Rényi [1964] points out, the idea of taking conditional probability as primitive seems to go back as far as Keynes [1921]. CP1–3 are essentially due to Rényi [1955]. 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. Starting from the work of de Finetti, there has been a general study of *coherent conditional probabilities.* A coherent conditional probability is essentially a cps that is not necessarily a Popper space, since it is defined on a set $\mathcal{F} \times \mathcal{F}'$ where $\mathcal{F}'$ does not have to be a subset of $\mathcal{F}$); see, for example, [Coletti and Scozzafava 2002] and the references therein. 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 2.2: 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 finitely additive 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 $(W, \mathcal{F})$. Again, if $\mathcal{F}$ is a σ-algebra, a superscript c is used to denote countable additivity, so $LPS^c(W, \mathcal{F})$ 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. I write $\vec{\mu}(U) > 0$ if $\mu_\beta(U) > 0$ for some β. ∎

There is a sense in which $LPS(W, \mathcal{F})$ can capture a richer set of preferences than $Pop(W, \mathcal{F})$, even if we restrict to finite spaces W (so that countable additivity is not an issue). For example, suppose that $W = \{w_1, w_2\}$, $\mu_0(w_1) = \mu_0(w_2) = 1/2$, and $\mu_1(w_1) = 1$. The LPS $\vec{\mu} = (\mu_0, \mu_1)$ can be thought of describing the situation where w_1 is very slightly more likely than w_2. Thus, for example, if X_i is a bet that pays off 1 in state w_i and 0 in state w_{3-i}, then according to $\vec{\mu}$, X_1 should be (slightly) preferred to X_2, but for all $r > 1$, rX_2 is preferred to X_1. There is no CPS on $\{w_1, w_2\}$ that leads to these preferences

Note that, in this example, the support of μ_2 is a subset of that of μ_1. To obtain a bijection between LPS's and CPS's, we cannot allow much overlap between the supports of the measures that make an LPS. What counts as "much overlap" turns out to be a somewhat subtle. One way to formalize it was proposed by BBD. They defined a *lexicographic conditional probability space (LCPS)* to be an LPS such that, roughly speaking, the probability measures in the sequence have disjoint supports; more precisely, there exist sets $U_\beta \in \mathcal{F}$ such that $\mu_\beta(U_\beta) = 1$ and the sets U_β are pairwise disjoint for $\beta < \alpha$. One motivation for considering disjoint sets is to consider an agent who has a sequence of hypotheses $(h_0, h_1, \ldots)$ regarding how the world works. If the primary hypothesis h_0 is discarded, then the agent judges events according to h_1; if h_1 is discarded, then the agent uses h_2, and so on. Associated with hypothesis h_β is the probability measure μ_β. What would cause h_β to be discarded is observing an event U such that $\mu_\beta(U) = 0$. The set U_β is the support of the hypothesis h_β. In some cases, it seems reasonable to think of the supports of these hypotheses as disjoint. This leads to LCPS's.

BBD considered only finite spaces. When we move to infinite spaces, requiring disjointness of the supports of hypotheses may be too strong. Brandenburger, Friedenberg, and Keisler [2008] consider finite-length LPS's $\vec{\mu}$ that satisfy the property that there exist sets U_β (not necessarily disjoint) such that $\mu_\beta(U_\beta) = 1$ and $\mu_\beta(U_\gamma) = 0$ for $\gamma \neq \beta$. Call such an LPS an MSLPS (for *mutually singular LPS*). Let a *structured LPS (SLPS)* be an LPS $\vec{\mu}$ such that there exist sets $U_\beta \in \mathcal{F}$ such that $\mu_\beta(U_\beta) = 1$ and $\mu_\beta(U_\gamma) = 0$ for $\gamma > \beta$. Thus, in an SLPS, later hypotheses are given probability 0 according to the probability measure induced by earlier hypotheses, but earlier hypotheses do not necessarily get probability 0 according the later hypotheses. (Spohn [1986] also considered SLPS's; he called them *dimensionally well-ordered families of probability measures*.) Clearly every LCPS is an MSLPS, and every MSLPS is an SLPS. If α is countable and we require countable additivity (or if α is finite) then the notions are easily seen to coincide. Given an SLPS $\vec{\mu}$ with associated sets $U_\beta, \beta < \alpha$, define $U'_\beta = U_\beta - (\cup_{\gamma>\beta} U_\gamma)$. The sets U'_β are clearly pairwise disjoint elements of $\mathcal{F}$, and $\mu_\beta(U'_\beta) = 1$. However, in general, LCPS's are a strict subset of MSLPS's, and MSLPS's are a strict subsets of SLPS's, as the following two examples show.

Example 2.3: Consider a well-ordering of the interval $[0, 1]$, that is, a bijection 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 we can take $U_0 = [0, 1]$ and $U_\beta = \{r_\beta\}$ for $0 < \beta < \alpha$; note that $\mu_\alpha(U_\beta) = 0$ for $\beta > \alpha$. However, this SLPS is not equivalent to any MSLPS (and hence not to any LCPS); there is no set U'_0 such that $\mu_0(U'_0) = 1$ and U'_0 is disjoint from r_β for all β with $0 < \beta < \alpha$. ∎

Example 2.4: Suppose that $W = [0, 1] \times [0, 1]$. Again, consider a well-ordering on $[0, 1]$. Using the notation of Example 2.3, define $U_{0,\beta} = r_\beta \times [0, 1]$ and $U_{1,\beta} = [0, 1] \times \{r_\beta\}$. Define $\mu_{i,\beta}$ to be the Borel measure on $U_{i,\beta}$. Consider the LPS $(\mu_{0,0}, \mu_{0,1}, \ldots, \mu_{1,0}, \mu_{1,1}, \ldots)$. Clearly this is an MSLPS, but not an LCPS. ∎

The difference between LCPS's, MSLPS'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. Here it will play a more significant role.

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 $\vec{\mu}(U) > 0$, 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 $\vec{\mu}(U) = 0$.

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 $\mathbb{R}(\epsilon)$, is the smallest field generated by adding to the reals a single infinitesimal ϵ.[2] We can further restrict to non-Archimedean fields that are *elementary extensions* of the standard reals: they agree with the standard reals on all properties that can be expressed in a first-order language with a predicate N representing the natural numbers. For most of this paper, I use only the following properties of non-Archimedean fields:

1. If $\mathbb{R}^*$ is a non-Archimedean field, then for all $b \in \mathbb{R}^*$ 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 $\mathbb{R}^*$, a *nonstandard probability space (NPS) over* $(W, \mathcal{F})$ *(with range* $\mathbb{R}^*$*)* 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 $\mathcal{F}$ a nonnegative element of $\mathbb{R}^*$ such that $\mu(W) = 1$ and $\mu(U \cup V) = \mu(U) + \mu(V)$ if U and V are disjoint.[3]

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 $F_{S \rightarrow P}$ 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, $F_{S \rightarrow P}$ is a bijection. 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 $V \in \mathcal{F}'$, let β_V be the smallest index such $\mu_{\beta_V}(V) > 0$. Define $\mu(U \,|\, V) = \mu_{\beta_V}(U \,|\, V)$. I leave it to the reader to check that $(W, \mathcal{F}, \mathcal{F}', \mu)$ is a Popper space.

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

1. $\mathcal{F}'$ consists precisely of those events for which conditioning in the LPS is defined; that is, $\mathcal{F}' = \{U : \vec{\mu}(U) > 0\}$.

[2]The construction of $\mathbb{R}(\epsilon)$ apparently goes back to Robinson [1973]. It is reviewed by Hammond [1994, 1999] and Wilson [1995] (who calls $\mathbb{R}(\epsilon)$ the *extended reals*).

[3]Note that, unlike Hammond [1994, 1999], I do not restrict the range of probability measures to consist of ratios of polynomials in ϵ with nonnegative coefficients.

2. For $U \in \mathcal{F}'$, $\mu(\cdot \,|\, U) = \mu'(\cdot \,|\, U)$, 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 is clear that these two properties in fact characterize $F_{S\to P}$. Thus, $F_{S\to P}$ preserves the events on which conditioning is possible and the most significant term in the lexicographic probability.

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 a bijection 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 finite Popper spaces with arbitrary conditioning events. No new conceptual issues arise in doing this extension; I include it here only to be able to contrast it with the other results.

Let $SLPS(W,\mathcal{F})$ denote the set of LPS's over $(W,\mathcal{F})$; let $SLPS(W,\mathcal{F},\mathcal{F}')$ denote the set of LPS's $(W,\mathcal{F},\vec{\mu})$ such that $\vec{\mu}(U) > 0$ for all $U \in \mathcal{F}'$ (i.e., $\mu_\beta(U) > 0$ for some β); as usual, I use a superscript c to denote countable additivity, so, for example, $SLPS^c(W,\mathcal{F})$ denotes the set of countably additive SLPS's over $(W,\mathcal{F})$. Let $Pop(W,\mathcal{F},\mathcal{F}')$ denote the set of Popper spaces of the form $(W,\mathcal{F},\mathcal{F}')$ and let $Pop^c(W,\mathcal{F},\mathcal{F}')$ denote the set of Popper spaces of the form $(W,\mathcal{F},\mathcal{F}',\mu)$ where μ is countably additive.

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

Proof: It is immediate from the definition that if $(W,\mathcal{F},\vec{\mu}) \in SLPS(W,\mathcal{F},\mathcal{F}')$, then $F_{S\to P}(W,\mathcal{F},\vec{\mu}) \in Pop(W,\mathcal{F},\mathcal{F}')$. It is also straightforward to show that $F_{S\to P}$ is an injection (see the appendix for details). The work comes in showing that $F_{S\to P}$ is a surjection (or, equivalently, in constructing an inverse to $F_{S\to P}$). I sketch the main ideas of the argument here, leaving details to the appendix.

Given $\mu \in Pop(W,\mathcal{F},\mathcal{F}')$, the idea is to choose $k \le |W|$ and k disjoint sets $U_0,\ldots,U_k \in \mathcal{F}'$ appropriately such that $\mu_j = \mu \,|\, U_j$ for $j = 0,\ldots,k$ (i.e., $\mu_j(V) = \mu(V \,|\, U_j)$) and $F_{S\to P}(W,\mathcal{F},\vec{\mu}) = \mu$. Since the sets $U_0,\ldots,U_k$ are disjoint, $\vec{\mu}$ must be an SLPS. The difficulty lies in choosing $U_0,\ldots,U_k$ so that $\vec{\mu}(U) > 0$ iff $U \in \mathcal{F}'$. This is done as follows. Let U_0 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 $\mu(U \mid W) = 1$. Since $\mu(U_0 \mid W) > 0$, it follows that $U_0 \in \mathcal{F}'$. If $\overline{U}_0 \notin \mathcal{F}'$. then (because $\mathcal{F}'$ is closed under supersets in $\mathcal{F}$), no subset of $\overline{U}_0$ is in $\mathcal{F}'$. If $\overline{U}_0 \in \mathcal{F}'$, let U_1 be the smallest set in $\mathcal{F}$ such that $\mu(U_1 \mid \overline{U}_0) = 1$. Note that $U_1 \subseteq \overline{U}_0$ and that $U_1 \in \mathcal{F}'$. Continuing in this way, it is clear that there exists a $k \geq 0$ and a sequence of pairwise disjoint sets $U_0, U_1, \ldots, U_k$ such that (1) $U_i \in \mathcal{F}'$ for $i = 0, \ldots, k$, (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) $\overline{U_0 \cup \ldots \cup U_k} \notin \mathcal{F}'$. Condition (2) guarantees that 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). ∎

Corollary 3.2: *If W is finite, then $F_{S \to P}$ is a bijection from* $\mathrm{SLPS}(W, \mathcal{F})$ *to* $\mathrm{Pop}(W, \mathcal{F})$.

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 SLPS's to Popper spaces. However, as the following two examples show, if we do not require countable additivity, then it is not a bijection.

Example 3.3: (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 $I\!N$ (recall that a cofinite set is the complement of a finite set), 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 $(I\!N, \mathcal{F}, \mathcal{F}', \mu^1)$ is a Popper space. Note that μ^1 is not countably additive (since $\mu^1(\{i\} \mid I\!N) = 0$ for all i, although $\mu^1(I\!N \mid I\!N) = 1$). Suppose that there were some LPS $(I\!N, \mathcal{F}, \vec{\mu})$ 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 $n \in I\!N$. 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 probability measure μ_i on the natural numbers such that $\mu_i(n) = \mu_i(0) > 0$ for all $n \geq 0$. For if $\mu_i(0) > 1/N$, then $\mu_i(\{0, \ldots, N-1\}) > 1$, a contradiction. (See Example 4.8 for further discussion of this setup.) ∎

Example 3.4: Again, let $W = I\!N$, let $\mathcal{F}$ consist of the finite and cofinite subsets of $I\!N$, 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 $\max(V \cap U) = \max U$, 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 $(I\!N, \mathcal{F}, \mathcal{F}', \mu^2)$ is a Popper space. Suppose there were some LPS $(I\!N, \mathcal{F}, \vec{\mu})$ which was mapped by $F_{S \rightarrow P}$ 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, originally proved by Spohn [1986], shows, there are no such counterexamples if we restrict to countably additive SLPS's and countably additive Popper spaces.

Theorem 3.5: [Spohn 1986] *For all W, the map $F_{S \rightarrow P}$ is a bijection from* $\mathrm{SLPS}^c(W, \mathcal{F}, \mathcal{F}')$ *to* $\mathrm{Pop}^c(W, \mathcal{F}, \mathcal{F}')$.

Proof: Again, the difficulty comes in showing that $F_{S \rightarrow P}$ is onto. Given a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$, I again construct sets $U_0, U_1, \ldots$ and an LPS $\vec{\mu}$ such that $\mu_\beta(V) = \mu(V \mid U_\beta)$, and show that $F_{S \rightarrow P}(W, \mathcal{F}, \vec{\mu}) = (W, \mathcal{F}, \mathcal{F}', \mu)$. However, now a completely different construction 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_0. There may no longer be a smallest set U_0 such that $\mu(U_0) = 1$. Consider, for example, the interval $[0, 1]$ with Borel measure. There is clearly no smallest subset U of $[0, 1]$ such that $\mu(U) = 1$. The details can be found in the appendix. ∎

Corollary 3.6: *For all W, the map $F_{S \rightarrow P}$ is a bijection from* $\mathrm{SLPS}^c(W, \mathcal{F})$ *to* $\mathrm{Pop}^c(W, \mathcal{F})$.

It is important in Corollary 3.6 that we consider SLPS's and not MSLPS's or LCPS's. $F_{S \rightarrow P}$ is in fact not a bijection from MSLPS's or LCPS's to Popper spaces.

Example 3.7: Consider the Popper space $([0, 1], \mathcal{F}, \mathcal{F}', \mu)$ which is the image under $F_{S \rightarrow P}$ of the SLPS constructed in Example 2.3. It is easy to see that this Popper space cannot be the image under $F_{S \rightarrow P}$ of some MSPLS (and hence not of some LCPS either). ∎

3.3 Treelike CPS's

One of the requirements in a Popper space is that $\mathcal{F}'$ be closed under supersets in $\mathcal{F}$. If we think of $\mathcal{F}'$ as consisting of all sets on which conditioning is possible, this makes sense; if we can condition on a set U, we should be able to consider on a superset V of U. But if we think of $\mathcal{F}'$ as representing all the possible evidence that can be obtained (and thus, the set of events on which an agent must be be able to condition, so as to update her beliefs), there is no reason that $\mathcal{F}'$ should be closed under supersets; nor, for that matter, is it necessarily the case that if $U \in \mathcal{F}'$ and $\mu(V \mid U) \neq 0$, then $V \cap U \in \mathcal{F}'$. In general, a cps where $\mathcal{F}'$ does not have these properties cannot be represented by an LPS, as the following example shows.

Example 3.8: Let $W = \{w_1, w_2, w_3, w_4\}$, let $\mathcal{F}$ consist of all subsets of W, and let $\mathcal{F}'$ consist of all the 2-element subsets of W. Clearly $\mathcal{F}'$ is not closed under supersets. Define μ on $\mathcal{F} \times \mathcal{F}'$ such that $\mu(w_1 \mid \{w_1, w_3\}) = \mu(w_4 \mid \{w_2, w_4\}) = 1/3$, and $\mu(w_1 \mid \{w_1, w_2\}) = \mu(w_4 \mid \{w_3, w_4\}) = 1/2$, and CP1 and CP2 hold. This is easily seen to determine μ. Moreover, μ vaciously satisfies CP3, since there do not exist distinct sets U and X in $\mathcal{F}'$ such that $U \subseteq X$. It is easy to show that there is no unconditional probability μ^* on W such that $\mu^*(U \mid V) = \mu(U \mid V)$ for all pairs $(U, V) \in \mathcal{F} \times \mathcal{F}'$ such that $\mu^*(V) > 0$ (where, for μ^*, the conditional probability is defined in the standard way).[4] It easily follows that there is no LPS $\vec{\mu}$ such that $\vec{\mu}(U \mid V) = \mu(U \mid V)$ for all $(U, V) \in \mathcal{F} \times \mathcal{F}'$ (since otherwise μ_0 would agree with μ on all pairs $(U, V) \in \mathcal{F} \times \mathcal{F}'$ such that $\mu(V) > 0$). Had $\mathcal{F}'$ been closed under supersets, it would have included W. It is easy to see that it is impossible to extend μ to $\mathcal{F} \times (\mathcal{F}' \cup \{W\})$ so that CP3 holds. ∎

In the game-theory literature, Battigalli and Siniscalchi [2002] use conditional probability measures to model players' beliefs about other players' strategies in extensive-form games where agents have perfect recall. The conditioning events are essentially information sets; which can be thought of as representing the possible evidence that an agent can obtain in a game. Thus, the cps's they consider are not necessarily Popper spaces, for the reasons described above. Nevertheless, the conditioning events considered by Battigalli and Sinischalchi satisfy certain properties that prevent an analogue of Example 3.8 from holding. I now make this precise.

Formally, I assume that there is a one-to-one correspondence between the sets in $\mathcal{F}'$ and the information sets of some fixed player i. For each set $U \in \mathcal{F}'$, there is a unique information set I_U for player i such that U consists of all the strategy profiles that reach I_U. With this identification, it is immediate that we can organize the sets in $\mathcal{F}'$ into a forest (i.e., a collection of trees), with the same "reachability" structure as that of the information sets in the game tree. The topmost sets in the forest are the ones corresponding to the topmost information sets for player i in the game tree. There

[4] This example is closely related to examples of conditional probabilities for which there is no common prior; see, for example, [Halpern 2002, Example 2.2].

may be several such topmost information sets if nature or some player j other than i makes the first move in the game. (That is why we have a forest, rather than a tree.) The immediate successors of a set U are the sets of strategy profiles corresponding to information sets for player i reached immediately after I_U. Because agents have perfect recall, the conditioning events $\mathcal{F}'$ have the following properties:

T1. $\mathcal{F}'$ is countable.

T2. The elements of $\mathcal{F}'$ can be organized as a forest (i.e., a collection of trees) where, for each $U \in \mathcal{F}'$, if there is an edge from U to some $U' \in \mathcal{F}'$, then $U' \subseteq U$, all the immediate successors of U are disjoint, and U is the union of its immediate successors.

T3. The topmost nodes in each tree of the forest form a partition of W.

Say that a set $\mathcal{F}'$ is *treelike* if it satisfies T1–3. It follows from T2 and T3 that, for any sets U and U' in a treelike set $\mathcal{F}'$, either $U \subseteq U'$ (if U is a descendant of U' in some tree), $U' \subseteq U$ (if U' is a descendant of U), or U and U' are disjoint (if neither is a descendant of the other). If $\mathcal{F}'$ is treelike, let $T^c(W, \mathcal{F}, \mathcal{F}')$ consist of all countably additive cps's defined on $\mathcal{F} \times \mathcal{F}'$. I abuse notation in the next result, viewing $F_{S\rightarrow P}$ as a mapping from $SLPS^c(W, \mathcal{F}, \mathcal{F}')$ to $T^c(W, \mathcal{F}, \mathcal{F}')$.

Proposition 3.9: *The map $F_{S\rightarrow P}$ is a surjection from* $\mathrm{SLPS}^c(W, \mathcal{F}, \mathcal{F}')$ *onto* $T^c(W, \mathcal{F}, \mathcal{F}')$.

Since $\mathcal{F}'$ is countable, every SLPS in $SLPS^c(W, \mathcal{F}, \mathcal{F}')$ must have at most countable length. Thus, there is no distinction between SLPS's, LCPS's, and MSPLS's in this case. (Indeed, in the proof of Proposition 3.9, the LPS constructed to demonstrate the surjection is an LCPS.) Note that we cannot hope to get a bijection here, even if W is finite. For example, suppose that $W = \{w_1, w_2\}$, $\mathcal{F} = 2^W$, and $\mathcal{F}' = \{\{w_1\}, \{w_2\}\}$. $\mathcal{F}'$ is clearly treelike, and there is a unique cps μ on $(W, \mathcal{F}, \mathcal{F}')$. $F_{S\rightarrow P}$ maps every SLPS in $SLPS(W, \mathcal{F}, F')$ to μ, but is clearly not a bijection. (This example also shows that we do not get a bijection by considering MSLPS's or LCPS's either.)

3.4 Related Work

It is interesting to contrast these results to those of Rényi [1956] and van Fraassen [1976]. Rényi 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 an algebra of subsets of W, $\mathcal{F}'$ is a subset of $\mathcal{F}$ closed under finite unions (but not necessarily closed under supersets in $\mathcal{F}$), 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 systems 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 a bijection from dimensionally ordered systems to cps's where the set $\mathcal{F}'$ is closed under finite unions. (Császár [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 3.3 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 3.4. This corresponds to the dimensionally ordered system $\{\mu_\beta : \beta \in I\!N \cup \{\infty\}\}$, 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.

Krauss [1968] restricts to Popper algebras of the form $\mathcal{F} \times (\mathcal{F} - \{\emptyset\})$; this allows him to simplify and generalize Rényi's analysis. Interestingly, he also proves a representation theorem in the spirit of Rényi's that involves nonstandard probability.

Van Fraassen [1976] proves a result whose assumptions are somewhat closer to Theorem 3.5. 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 $F_{O\to C}$ 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 $F_{O\to C}$ is a bijection 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 $F_{O\to C}$ is also a bijection from countably additive ordinal families to countably additive Popper spaces.

Spohn's result, Theorem 3.5, can be viewed as a strengthening of van Fraassen's result in the countably additive case, since for Theorem 3.5 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 3.3 corresponds to the ordinal family $\{(W_n, \mathcal{F}_n, \mu_n) : n \le \omega\}$ 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 3.4 is represented in a similar way, using the ordinal family $\{(W_n, \mathcal{F}_n, \mu'_n) : n \le \omega\}$, 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. I now 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 random variables 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 (real valued) random variables (according to their expected value) is the same. I consider only random variables with countable range. This restriction both makes the exposition simpler and avoids having to define, for example, integration with respect to an NPS. 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 real-valued random variables X and Y measurable with respect to $\mathcal{F}$, $E_{\nu_1}(X) \leq E_{\nu_1}(Y)$ iff $E_{\nu_2}(X) \leq E_{\nu_2}(Y)$. (If X has countable range, which is the only case I consider here, then X is measurable with respect to F iff $\{w : X(w) = x\} \in \mathcal{F}$ for all x in the range of X.)[5] ∎

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 $\mu_j(U) > 0$.*

[5] As pointed out by Adam Brandenburger and Eddie Dekel, this notion of equivalence is essentially the same as one implicitly used by BBD. They work with preference orders on Anscombe-Aumann acts [Anscombe and Aumann 1963], that is, functions from states to probability measures on prizes. Fix a utility function u on prizes. Then take $\nu_1 \sim_u \nu_2$ if the preference order on acts generated by ν_1 and u is the same as that generated by ν_2 and u. It is not hard to show that this notion of equivalence is independent of the choice of utility function; if u and u' are two utility functions on prizes, then $\nu_1 \sim_u \nu_2$ iff $\nu_1 \sim_{u'} \nu_2$. Moreover, $\nu_1 \sim_u \nu_2$ iff $\nu_1 \approx \nu_2$. The advantage of the notion of equivalence used here is that it is defined without the overhead of preference orders on acts.

As the next result shows, for SLPS's, the $\approx$-equivalence classes are singletons, even if the set of worlds is infinite. (This is not true for LPS's in general. For example, $(\mu, \mu) \approx (\mu)$.) This can be viewed as providing more motivation for the use of SLPS's.

Proposition 4.3: *If $\vec{\mu}, \vec{\mu}' \in \mathrm{SLPS}(W, \mathcal{F})$, then $\vec{\mu} \approx \vec{\mu}'$ iff $\vec{\mu} = \vec{\mu}'$.*

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 4.4: *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 bijection 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 the nonstandard model $I\!R(\epsilon)$ 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 4.5: *If W is finite, then $F_{L \to N}$ is a bijection 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: It is easy to check that if $\vec{\mu} = (\mu_0, \ldots, \mu_k)$, then $\vec{\mu} \approx (1 - \epsilon - \cdots - \epsilon^k)\mu_0 + \epsilon\mu_1 + \cdots + \epsilon^k \mu_k$ (see Lemma A.7 in the appendix for a formal proof). It follows that $F_{L \to N}$ is an injection from $LPS(W, \mathcal{F})/\!\approx$ to $NPS(W, \mathcal{F})/\!\approx$. To show that $F_{L \to N}$ is a surjection, we must essentially construct an inverse map; that is, given an NPS $(W, \mathcal{F}, \nu)$ where W is finite, we must find an LPS $\vec{\mu}$ such that $\vec{\mu} \approx \nu$. The idea is to find a finite collection $\mu_0, \ldots, \mu_k$ of (standard) probability measures, where $k \leq |W|$, and nonnegative nonstandard reals $\epsilon_0, \ldots, \epsilon_k$ such that $st(\epsilon_{i+1}/\epsilon_i) = 0$ and $\nu = \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$. A straightforward argument then shows that $\nu \approx \vec{\mu}$ and $F_{L \to N}([\vec{\mu}]) = [\nu]$. I leave details to the appendix. ∎

BBD [1991a] also relate nonstandard probability measures and LPS's under the assumption that the state space is finite, but there are some significant technical differences between the way they relate them and the approach taken here. BBD prove representation theorems essentially showing that a preference order 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 4.5 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 4.5 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 4.4 can be used to provide a bound on the length of the minimal-length LPS's in an equivalence class.

Proposition 4.6: *Every LPS over $(W, \mathcal{F})$ is equivalent to an LPS over $(W, \mathcal{F})$ of length at most $|\mathcal{F}|$.*

The first step in relating LPS's to NPS's is to show that, just as in the finite case, for every LPS $(\mu_\beta : \beta < \alpha)$ of length α, there is an equivalent NPS ν. The idea will be to set $\nu = (1 - \sum_{0<\beta<\alpha} \epsilon^{n_\beta}) + \sum_{0<\beta<\alpha} \epsilon_{n_\beta}\mu_\beta$. In the finite case, we could take $n_\beta = \beta$. This worked because each β was finite, and the field $I\!R(\epsilon)$ includes ϵ^j for each integer j. But now, since α may be greater than ω, we cannot just take $n_\beta = \beta$. To get this idea to work in the infinite setting, I consider a *nonstandard* model of the integers, which includes an "integer" corresponding to all the ordinals less than α. I then construct a field that includes ϵ^{n_α} even for these nonstandard integers n_α.

A *nonstandard model of the integers* is 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 I^α of the integers I^α that includes elements n_β, $\beta < \alpha$, such that $n_j = j$ for $j < \omega$ and $n_\beta < n_{\beta'}$ if $\beta < \beta'$. (Note that since I^α satisfies all the properties of the integers, it follows that if $n_\beta < n_{\beta'}$, then $n_{\beta'} - n_\beta \geq 1$, a fact that will be useful later.) The compactness theorem says that, given a collection of 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 $\mathbf{n}_i = i$ for $i < \omega$ 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. Thus, for each ordinal α, there is a model I^α with the required properties.

Given α, I now construct a field $I\!R(I^\alpha)$ that includes ϵ^n for each "integer" $n \in I^\alpha$. To explain the construction, it is best to first consider $I\!R(\epsilon)$ in a little more detail. Since $I\!R(\epsilon)$ is a field, once it includes ϵ, it must include $p(\epsilon)$, where p is a polynomial with real coefficients. To ensure the every nonzero element of $I\!R(\epsilon)$ has an inverse, we need not just finite polynomials in ϵ, but *infinite* polynomials in ϵ. The inverse of a polynomial in ϵ can then be computer using standard "formal" division of polynomials. Moreover, the leading coefficient of the polynomial can be negative. Thus, the inverse of ϵ^3 is, not surprisingly, ϵ^{-3}; the inverse of $1 - \epsilon$ is $1 + \epsilon + \epsilon^2 + \ldots$.

The field $I\!R(I^\alpha)$ also includes polynomials in ϵ, but now the exponents are not just integers, but elements of I^α. Since a field is closed under multiplication, if it contains ϵ^{n_1} and ϵ^{n_2}, it must also include their product. Since I^α satisfies all the properties of the integers, if it includes n_1 and n_2, it also includes an element $n_1 + n_2$, and we can take $\epsilon^{n_1} \times \epsilon^{n_2} = \epsilon^{n_1+n_2}$. Formally, let $I\!R(I^\alpha)$ be the non-Archimedean model defined as follows: $I\!R(I^\alpha)$ consists of all polynomials of the form $\sum_{n \in J} r_n \epsilon^n$, where r_n is a standard real, ϵ is an infinitesimal, and J is a *well-founded* subset of I^α. (Recall that a set is well founded if it has no infinite descending sequence; thus, the set of integers is not well founded, since $\ldots - 3 < -2 < -1$ is an infinite descending sequence. The reason I require well foundedness will be clear shortly.) We can identify the standard real r with the polynomial $r\epsilon^0$.

The polynomials in $I\!R(I^\alpha)$ 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 $I\!R(I^\alpha)$. In particular, if p_1 and p_2 are two polynomials, N_1 is the set of exponents of p_1, and N_2 is the set of exponents of p_2, then the exponents of $p_1 + p_2$ lie in $N_1 \cup N_2$, while the exponents of $p_1 p_2$ lie in the set $N_3 = \{n_1 + n_2 : n_1 \in N_1, n_2 \in N_2\}$. 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 $n + n' = n_1 + n_2$, so the coefficient of $\epsilon^{n_1+n_2}$ in $p_1 p_2$ is well defined. 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. This step is where the well foundedness comes in. The formal division process cannot be applied to a polynomial with coefficients that are not well founded, such as $\cdots + \epsilon^{-3} + \epsilon^{-2} + \epsilon^{-1}$. An element of $I\!R(I^\alpha)$ is *positive* if its leading coefficient is positive. Define an order $\leq$ on $I\!R(I^\alpha)$ by taking $a \leq b$ if $b - a$ is positive. With these definitions, $I\!R(I^\alpha)$ is a non-Archimedean field.

Given $(W, \mathcal{F})$, let α be the minimal ordinal whose cardinality is greater than or equal to $|\mathcal{F}|$. By construction, I^α has elements n_β for all $\beta < \alpha$ such that $n_i = i$ for $i < \omega$ and $n_\beta < n_{\beta'}$ if $\beta < \beta' < \alpha$. I now define a map $F_{L\rightarrow N}$ from $LPS(W, \mathcal{F})/\approx$ to $NPS(W, \mathcal{F})/\approx$ just as suggested earlier. In more detail, given an equivalence class $[\vec{\mu}] \in LPS(W, \mathcal{F})$, by Proposition 4.6, there exists $\vec{\mu}' \in [\vec{\mu}]$ such that $\vec{\mu}'$ has length $\alpha' \leq \alpha$. Let $\nu = (1 - \sum_{0<\beta<\alpha} \epsilon^{n_\beta})\mu_0 + \sum_{0<\beta<\alpha} \epsilon_{n_\beta}\mu'_\beta$. By definition, $\sum_{0<\beta<\alpha} \epsilon^{n_\beta} \in I\!R(I^\alpha)$ (the set of exponents is well ordered since the ordinals are well ordered), hence so is $(1 - \sum_{0<\beta<\alpha} \epsilon^{n_\beta})$. The elements ϵ^{n_β} for $\beta \leq \alpha$ are also all in $I\!R(I^\alpha)$. It easily follows that ν is nonstandard probability measure over the field $I\!R(I^\alpha)$. As observed earlier, if $\beta' < \beta$, then $\beta - \beta' \geq 1$, so $\epsilon^{n'_\beta}$ is infinitesimally smaller than ϵ^{n_β}. Arguments essentially identical to those of Lemma A.7 in the appendix can be used to show that $\nu \approx \vec{\mu}'$. Define $F_{L\rightarrow N}[\vec{\mu}] = [\nu]$. The following result is immediate.

Theorem 4.7: *$F_{L\rightarrow N}$ 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? The technique for finding an equivalent LPS used in the finite case fails. There is no obvious way to find a well-ordered sequence of standard probability measures $\mu_0, \mu_1, \ldots$ and a sequence of nonnegative nonstandard reals $\epsilon_0, \epsilon_1, \ldots$ such that $st(\epsilon_{\beta+1}/\epsilon_\beta) = 0$ and $\nu = \epsilon_0\mu_0 + \epsilon_1\mu_1 + \cdots$. As the following example shows, this is not an accident. There exists NPSs that are not equivalent to any LPS.

Example 4.8: As in Example 3.3, 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 $\nu^1(U) = 1 - |\overline{U}|\epsilon$ if U is cofinite (as usual, $\overline{U}$ denotes the complement of U, which in this case is finite). This is clearly an NPS, and it corresponds to the cps μ^1 of Example 3.3, in the sense that $st(\nu^1(V \mid U)) = \mu^1(V \mid U)$ for all $V \in \mathcal{F}$, $U \in \mathcal{F}'$. Just as in Example 3.3, it can be shown that there is no LPS $\vec{\mu}$ such that $\nu^1 \approx \vec{\mu}$.

To see the potential relevance of this setup, suppose that a natural number is chosen at random and, intuitively, all numbers are equally likely to be chosen. An agent may place a bet on the number being in a finite or cofinite set. Intuitively, the agent should prefer a bet on a set with larger cardinality. More precisely, if U_1 and U_2 are two sets in the algebra, the agent should prefer a bet on U_1 over a bet on U_2 iff (a) U_1 and U_2 are both cofinite and the complement of U_1 has smaller cardinality than that of U_2, (b) U_1 is cofinite and U_2 is finite, or (c) U_1 and U_2 are both finite, and U_1 has larger cardinality than U_2. These preferences on acts or bets should translate to statements of likelihood. The NPS captures these preferences directly; they cannot be captured in an LPS. The cps of Example 3.3 captures (b) directly, and (c) indirectly: when conditioning on any finite set that contains $U_1 \cup U_2$, the probability of U_1 will be higher than that of U_2. ∎

4.3 Countably additive nonstandard probability measures

Do things get any better if countable additivity is required? To answer this question, I 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.) Note that here and elsewhere I use the letters a and b (possibly with subscripts) to denote (standard) reals, and ϵ to denote an infinitesimal. 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.[6] 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 Theorem 4.7 can be applied with no change in proof to show the following.

Theorem 4.9: *$F_{L \rightarrow N}$ is an injection from* $\mathrm{LPS}^c(W, \mathcal{F})/\approx$ *to* $\mathrm{NPS}^c(W, \mathcal{F})/\approx$.

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 4.10: 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

[6] 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.

taking $\nu(w_j) = a_j + b_j\epsilon$, where $a_j = 1/2^j$, $b_{2j-1} = 1/2^{j-1}$, and $b_{2j} = -1/2^{j-1}$, 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. These likelihoods correspond to preferences. For example, an agent should prefer a bet that gives a payoff of 1 if w_2 occurs and 0 otherwise to a bet that gives a payoff of 4 if w_4 occurs and 0 otherwise. 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 4.10 is that the ratio between a_j and b_j in the definition of ν (i.e., the ratio between the "standard part" of $\nu(w_j)$ and the "infinitesimal part" of $\nu(w_j)$) goes to zero. 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.

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. This is arguably the most natural map; for example, it is easy to check that $F_{N\to P} \circ F_{S\to N} = F_{S\to P}$, where $F_{S\to N}$ is the restriction of $F_{L\to N}$ to SLPSs. (Note that $F_{L\to N}$ is well-defined on SLPS's, since if $\vec{\mu}$ is an SLPS, by Proposition 4.3, $[\vec{\mu}] = \{\vec{\mu}\}$.)

We might hope that $F_{N\to P}$ is a bijection from $NPS(W, \mathcal{F})/\approx$ to $Pop(W, \mathcal{F})$. As I show shortly, it is not. To understand $F_{L\to N}$ better, 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 $\nu_1(U) \neq 0$. Thus, $\simeq$ essentially says that infinitesimal differences between conditional probabilities do not count. 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 from $NPS^c/\simeq$ to $Pop^c(W, \mathcal{F})$.

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

Proof: It is easy to see that $F_{N\to P}$ is an injection. In the countable case, the inverse map can be defined using earlier results. If $(W, \mathcal{F}, \mathcal{F}', \mu) \in Pop^c(W, \mathcal{F})$, by Theorem 3.5, there is a countably additive SLPS $\vec{\mu}'$ such that $F_{S\to P}((W, \mathcal{F}, \vec{\mu}')) = (W, \mathcal{F}, \mathcal{F}', \mu)$. By

Theorem 4.7, there is some $(W, \mathcal{F}, \nu) \in NPS^c(W, \mathcal{F})$ such that $\nu \approx \vec{\mu}'$. It is not hard to show that $F_{N \to P}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$; see the appendix for details. Showing that $F_{N \to P}$ is a surjection in the finitely additive case requires more work; again, see the appendix for details. ∎

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* $\nu_1 \approx \nu_2$ *then* $\nu_1 \simeq \nu_2$.

The converse of Proposition 5.2 does not hold in general. As a result, 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, in finite spaces, Theorem 3.1 shows that Popper spaces are equivalent to SLPS's, while Theorem 4.5 shows that $LPS(W, \mathcal{F})/\approx$ is equivalent to $NPS(W, \mathcal{F})/\approx$. By Proposition 4.3, $SLPS(W, \mathcal{F})/\approx$ is essentially identical to $SLPS(W, \mathcal{F})$ (all the equivalence classes in $SLPS(W, \mathcal{F})/\approx$ are singletons), so in finite spaces, the gap in expressive power between Popper spaces and NPS's essentially amounts to the gap between $SLPS(W, \mathcal{F})$ and $LPS(W, \mathcal{F})/\approx$. This gap is nontrivial. For example, there is no SLPS equivalent to the LPS (μ_1, μ_2) that represents the NPS in Example 5.3.

6 Independence

The notion of independence is fundamental. As I show in this section, the results of the previous sections sheds light on various notions of independence considered in the literature for LPS's and (variants of) cps's. I first consider independence for events and then independence for random variables. I then relate my definitions to those of BBD, Hammond, and Kohlberg and Reny [1997].

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.[7] 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 $\nu(U) \neq 0$ implies that $\nu(V \mid U) - \nu(V)$ 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 $V' = \{w_4\}$; as before, let $U = \{w_2, w_4\}$. It is easy to check that $st(\nu_1(V' \mid U)) = st(\nu_1(V')) = 0$, 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. Similarly, U can be approximately independent of V without $\overline{U}$ being approximately independent of V. For example, it is easy to check that $\overline{V}'$ is approximately independent of U with respect to ν_1, although V' is not.

A straightforward argument shows that U is approximately independent of V with respect to ν iff $\nu(U) \neq 0$ implies $st((\nu(V \cap U) - \nu(V) \times \nu(U))/\nu(U)) = 0$, while V is

[7] 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 $U, V \in \mathcal{F}$. I continue to be sloppy at times, reverting to more careful notation when necessary.

approximately independent of U with respect to ν iff the same statement holds with the roles of V and U reversed. 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 $\nu(U \cap V') \neq 0$ implies $\nu(V \mid U \cap V') = \nu(V \mid V')$. 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 (approximate) independence.

The following proposition shows that, although independence is not preserved by equivalence, approximate independence 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 $U \cap V' \notin \mathcal{F}'$, 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 bijection $F_{N \to P}$. 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? Interestingly, neither BBD nor Hammond define independence directly for LPS's. However, they do give definitions in terms of NPS's that can be applied to equivalent LPS's; indeed, BBD [1991b] do just this (see the discussion of BBD strong independence below).

I now consider independence for random variables. If X is a random variable on W, let $\mathcal{V}(X)$ denote range (set of possible values) of random variable X; that is, $\mathcal{V}(X) = \{X(w) : w \in W\}$. Recall that I am assuming that all random variables have countable range. Random variable X is independent of Y with respect to a standard probability measure μ if the event $X = x$ is independent of the event $Y = y$ with respect to μ, for all $x \in \mathcal{V}(X)$ and $y \in \mathcal{V}(Y)$. By analogy, for nonstandard probability measures, following Kohlberg and Reny [1997], define X and Y to be *weakly independent* with respect to ν if $X = x$ is approximately independent of $Y = y$ and $Y = y$ is approximately independent of $X = x$ with respect to ν for all $x \in \mathcal{V}(X)$ and $y \in \mathcal{V}(Y)$.[8]

For standard probability measures, 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 6.4: 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:

	$Y = 1$	$Y = 2$
$X = 1$	$1 - 3\epsilon - 3\epsilon^2$	ϵ
$X = 2$	ϵ	ϵ^2
$X = 3$	ϵ	$2\epsilon^2$

It is easy to check that X and Y are weakly independent 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* $\{Y_1, \ldots, Y_n\}$ *with respect to* ν if $X \in U_1$ is approximately independent of $(Y_1 \in V_1) \cap \ldots \cap (Y_n \in V_n)$ conditional on $(Y_1 \in V_1') \cap \ldots \cap (Y_n \in V_n')$ with respect to ν for all $U_1 \subseteq \mathcal{V}(X)$, $V_i, V_i' \subseteq \mathcal{V}(Y_i)$, and $i = 1, \ldots, n$. $X_1, \ldots, X_n$ are *approximately independent with respect to* ν if X_i

[8] Kohlberg and Reny's definition of weak independence also requires that the joint range of X and Y be the product of the individual ranges. That is, for X and Y to be weakly independent, it must be the case that for all $x \in \mathcal{V}(X)$ and $y \in \mathcal{V}(Y)$, there exists some $w \in W$ such that $X(w) = x$ and $Y(w) = y$. Of course, this requirement could also be added to the definition I am proposing here; adding it would not affect any of the results of this paper.

is approximately independent of $\{X_1, \ldots, X_n\} - \{X_i\}$ with respect to ν for $i = 1, \ldots, n$. I leave to the reader the obvious extensions to conditional independence and the analogues of this definition for Popper spaces and LPS's.

As I said, BBD consider three notions of independence for random variables. One is 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 probability measure ν such that $X_1, \ldots, X_n$ are approximately independent with respect to ν [Battigalli and Veronesi 1996]. A second notion they consider is 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.

The third notion of independence that BBD consider is the strongest. BBD [1991b] define $X_1, \ldots, X_n$ to be strongly independent with respect to an LPS $\vec{\mu}$ if they are independent (in the usual sense) with respect to an NPS ν such that $\mu \approx \nu$.[9] Moreover, they give a characterization of this notion of strong independence, which I henceforth call *BBD strong independence*, to distinguish it from the KR notion of strong independence that I discuss shortly. Given a tuple $\vec{r} = (r^0, \ldots, r^{k-1})$ of vectors of reals in $(0,1)^k$ and a finite LPS $\vec{\mu} = (\mu^0, \ldots, \mu^k)$, let $\vec{\mu} \Box \vec{r}$ be the (standard) probability measure

$$(1 - r^0)\mu^0 + r^0[(1 - r^1)\mu^1 + r^1[(1 - r^2)\mu^2 + r^2[\cdots + r^{k-2}[(1 - r^{k-1})\mu^{k-1} + r^{k-1}\mu^k] \ldots]]].$$

Note that $\vec{\mu} \Box \vec{r}$ is defined only if $\vec{\mu}$ is finite. Thus, in discussing BBD strong independence, I restrict to finite LPS's. In addition, for technical reasons that will become clear in the proof of Theorem 6.5, I consider only random variables with finite range, which is what BBD do as well. BBD [1991b, p. 90] claim without proof that "it is straightforward to show" that $X_1, \ldots, X_n$ are BBD strongly independent with respect to $\vec{\mu}$ iff there is a sequence $\vec{r}^j$, $j = 1, 2, \ldots$ of vectors in $(0,1)^k$ such that $\vec{r}^j \to (0, \ldots, 0)$ as $j \to \infty$, and $X_1, \ldots, X_n$ are independent with respect to $\vec{\mu} \Box \vec{r}^j$ for $j = 1, 2, 3, \ldots$. I can prove this result only if the NPS ν such that $\vec{\mu} \approx \nu$ and $X_1, \ldots, X_n$ are independent with respect to ν has a range that is an elementary extension of the reals (and thus has the same first-order properties as the reals).

Theorem 6.5: *There exists an NPS ν whose range is an elementary extension of the reals such that $\vec{\mu} \approx \nu$ and $X_1, \ldots, X_n$ are independent with respect to ν iff there exists a sequence $\vec{r}^j$, $j = 1, 2, \ldots$ of vectors in $(0,1)^k$ such that $\vec{r}^j \to (0, \ldots, 0)$ as $j \to \infty$, and $X_1, \ldots, X_n$ are independent with respect to $\vec{\mu} \Box \vec{r}^j$ for $j = 1, 2, 3, \ldots$.*

[9] In [Blume, Brandenburger, and Dekel 1991b], BBD say that this definition of strong independence is given in [Blume, Brandenburger, and Dekel 1991a]. However, the definition appears to be given only in terms of NPS's in [Blume, Brandenburger, and Dekel 1991a].

I do not know if this result holds without requiring that ν be an elementary extension of the reals.

Kohlberg and Reny [1997] define a notion of strong independence with respect to what they call *relative probability spaces*, which are closely related to Popper spaces of the form $(W, 2^W, 2^W - \{\emptyset\}, \mu)$, where all subsets of W are measurable and it is possible to condition on all nonempty sets. Their definition is similar in spirit to the characterization of BBD strong independence given in Theorem 6.5. For ease of exposition, I recast their definition in terms of Popper spaces. $X_1, \ldots, X_n$ are *KR-strongly independent* with respect to the Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$, where $\mathcal{F}'$ includes all events of the form $X_i = x$ for $x \in \mathcal{V}(X_i)$, if there exist a sequence of standard probability measures $\mu_1, \mu_2, \ldots$ such that $\mu_j \rightarrow \mu$, and for all $j = 1, 2, 3, \ldots$, $\mu_j(U) > 0$ for $U \in \mathcal{F}'$ and $X_1, \ldots, X_n$ are independent with respect to μ_j. As Kohlberg and Reny show, KR-strong independence implies approximate independence[10] and is, in general, strictly stronger.

The following theorem characterizes KR strong independence in terms of NPS's.

Theorem 6.6: *$X_1, \ldots, X_n$ are KR-strongly independent with respect to the Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$ iff there exists an NPS $(W, \mathcal{F}, \nu)$ such that $F_{N \rightarrow P}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$ and $X_1, \ldots, X_n$ are independent with respect to $(W, \mathcal{F}, \nu)$.*

It follows from the proof that we can require the range of ν to be a nonelementary extension of the reals, but this is not necessary.

Kohlberg and Reny show that their notions of weak and strong independence can be used to characterize Kreps and Wilson's [1982] notion of sequential equilibrium. BBD [1991b] use their notion of strong independence in their characterization of perfect equilibrium and proper equilibrium for games with more than two players. Finally, Battigali [Battigalli 1996] uses approximate independence (or, equivalently, independence in cps's) to characterize sequential equilibrium.

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. It would be of interest to get a natural characterization of those NPS's that are equivalent to Popper spaces and LPS's; this remains an open problem. 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 3.3 and 3.4 show. Since all of these approaches to representing uncertainty have been using in characterizing solution concepts in extensive-form games and notions

[10] They actually show only that it implies weak independence, but the same argument shows that it implies approximate independence.

of admissibility, the results here suggest that it is worth considering the extent to which these results depend on the particular representation used.

It is worth stressing here that this notion of equivalence depends on the fact that I have been viewing cps's, LPS's, and NPS's as representations of uncertainty. But, as Asheim [2006] emphasizes, they can also be viewed as representations of conditional preferences. Example 5.3 shows that, even in finite spaces, NPS's and LPS's can express preferences that cps's cannot. However, as Asheim and Perea [2005] point out, in finite spaces, cps's can also represent conditional preferences that cannot be represented by LPS's and NPS's. See [Asheim 2006] for a detailed discussion of the expressive power of these representations with respect to conditional preferences.

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. In particular, it may be of interest to consider LPS's indexed by partially ordered and not necessarily well-founded sets, rather than just LPS's indexed by the ordinals. For example, Brandenburger, Friedenberg, and Keisler [2008] characterize n rounds of iterated deletion using finite LPS's, for any n. Rather than using a sequence of (finite) LPS's of different lengths to characterize (unbounded) iterated deletion, it seems that a result similar in spirit can be obtained using a single LPS indexed by the (positive and negative) integers.

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. There are two standard notions of belief that can be defined in LPS's. Say that U is a *certain belief* in LPS $\vec{\mu}$ of length α if $\mu_\beta(U) = 1$ for all $\beta < \alpha$; U is *weakly believed* if $\mu_0(U) = 1$. Brandenburger, Friedenberg, and Keisler [2008] defined a third notion of belief, intermediate between weak and strong belief, and provided an elegant decision-theoretic justification of it. According to their definition, an agent *assumes* U *in* $\vec{\mu}$ if there is some $\beta < \alpha$ such that (a) $\mu_{\beta'}(U) = 1$ for all $\beta' \leq \beta$, (b) $\mu_{\beta''}(U) = 0$ for all $\beta'' > \beta$, and (c) $U \subseteq \cup_{\beta' \leq \beta} Supp(\mu_{\beta'})$, where $Supp(\mu_{\beta'})$ denotes the support of the probability measure $\mu_{\beta'}$. (Condition (c) is unnecessary if W is finite, given Brandenburger, Friedenberg, and Keisler's assumption that $W = \cup_{\beta'} Supp(\mu_{\beta'})$.) There are straightforward analogues of certain belief and weak belief in Popper spaces. U is strongly believed in a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$ if $\mu(U \mid V) = 1$ for all $V \in \mathcal{F}'$; U is weakly believed if $\mu(U \mid V) = 1$ for all $V \in \mathcal{F}'$ such that $\mu(V) > 0$. Analogues of this notion of assumption have been considered elsewhere in the literature. Van Fraassen [1995] independently defined a notion of belief using Popper spaces; in a finite state space, an event is what van Fraassen calls a *belief core* iff it is assumed in the sense of Brandenburger, Friedenberg, and Keisler. Battigalli and Siniscalchi's [2002] notion of *strong belief* is also essentially equivalent. Assumption also corresponds to Stalnaker's [1998] notion of *absolutely robust belief* and Asheim and Søvik's [2005] notion of *robust belief.* Asheim and Søvik [2005] do a careful comparison of all these notions (and others).

It is easy to define analogues of certain and weak belief in NPS's: U is certain belief if $\nu(U) = 1$; U is weakly believed if $st(\nu(U)) = 1$. The results of this paper suggest that it may also be worth investigating an analogue of assumption in NPS's.

- Nonstandard utility: In this paper, while I have allowed probabilities to be lexicographically ordered or nonstandard, I have implicitly assumed that utilities are standard real numbers (since I have restricted to real-valued random variables). 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) and Hammond [1999] 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 $\nu(U_n) \to \nu(U_\infty) = 1/3$. 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 $\mu_1 = st(\nu)$, $\mu_2(w_1) = 0$, $\mu_2(w_2) = 2/3$, 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'' is in the support 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 $\nu_n \to \nu$, 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).

  Of course, continuity plays a significant role in standard axiomatizations of SEU, and is vital in proving the existence of a Nash equilibrium. None of the uses of continuity that I am familiar with have the specific form of this example, but I believe it is worth considering further the impact of this lack of continuity.

Acknowledgments: I'd like to thank Adam Brandenburger and Peter Hammond for a number of very enlightening discussions, Bob Stalnaker for pointing out Example 3.3, Brian Skyrms for pointing me to Hammond's work, Bas van Fraassen for pointing me to Spohn's work, Amanda Friedenberg for her careful reading of an earlier draft, her many useful comments, and for encouraging me to try to understand what my results had to say about Battigalli and Sinischalchi's work, and Horacio Arlo-Costa, Geir Asheim, Larry Blume, Adam Brandenburger, Eddie Dekel, and the anonymous reviewers for a number of useful comments on earlier drafts 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 and $(\mathcal{F}, \mathcal{F}')$, then $F_{S\to P}$ is a bijection from* $\mathrm{SLPS}(W, \mathcal{F}, \mathcal{F}')$ *to* $\mathrm{Pop}(W, \mathcal{F}, \mathcal{F}')$.

Proof: The first step is to show that $F_{S\to P}$ is an injection. If $\vec{\mu}, \vec{\mu}' \in SLPS(W, \mathcal{F}, \mathcal{F}')$ and $\vec{\mu} \neq \vec{\mu}'$, let $\mu = F_{S\to P}(\mathcal{W}, \mathcal{F}, \vec{\mu})$, and let $\mu' = F_{S\to P}(\mathcal{W}, \mathcal{F}, \vec{\mu}')$. Let i be the least index such that $\mu_i \neq \mu_i'$. There is some set U such that $\mu_i(U) \neq \mu_i'(U)$. Let U_i be the set such $\mu_i(U_i) = 1$ and $\mu_j(U_i) = 0$ for $j < i$; since $\vec{\mu}$ is an SLPS, such a set U_i exists. Similarly, let U_i' be such that $\mu_i'(U_i) = 1$ and $\mu_j'(U_i) = 0$ for $j < i$. Since $\mu_j = \mu_j'$ for all $j < i$, we must have $\mu_j(U_i \cup U_i') = \mu_j(U_i \cup U_i') = 0$ for all $j < i$. Clearly $\vec{\mu}(U_j \cup U_j') > 0$, so $U_j \cup U_j' \in \mathcal{F}'$. Moreover, $\mu(U \mid U_i \cup U_i') = \mu_i(U \mid U_i \cup U_i') = \mu_i(U)$. Similarly, $\mu'(U \mid U_i \cup U_i') = \mu_i'(U)$. Hence, $\mu \neq \mu'$.

To show that $F_{S\to P}$ is a surjection, given a cps μ, let $\vec{\mu} = (\mu_0, \ldots, \mu_k)$ be the LPS constructed in the main text. We must show that $F_{S\to P}(\vec{\mu}) = (W, \mathcal{F}, \mathcal{F}', \mu)$. Suppose that $F_{S\to P}(\vec{\mu}) = (W, \mathcal{F}, \mathcal{F}'', \mu')$. 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, $\mu(U_j \mid \overline{U_0 \cup \ldots \cup U_{j-1}}) = 1$. It easily follows that if $V \subseteq \overline{U_0 \cup \ldots \cup U_{j-1}}$ then

$$\mu(V \mid \overline{U_0 \cup \ldots \cup U_{j-1}}) = \mu(V \cap U_j \mid \overline{U_0 \cup \ldots \cup U_{j-1}}).$$

Thus, by CP3,

$$\mu(V \mid \overline{U_0 \cup \ldots \cup U_{j-1}}) = \mu(V \cap U_j \mid \overline{U_0 \cup \ldots \cup U_{j-1}}) = \mu(V \mid U_j) \times \mu(U_j \mid \overline{U_0 \cup \ldots \cup U_{j-1}}),$$

so

$$\mu(V \mid U_j) = \mu(V \mid \overline{U_0 \cup \ldots \cup U_{j-1}}). \qquad (1)$$

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

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

Although Theorem 3.5 was proved by Spohn [1986], I include a proof here as well, to make the paper self-contained.

Theorem 3.5: *For all W, the map $F_{S \to P}$ is a bijection from* $\mathrm{SLPS}^c(W, \mathcal{F}, \mathcal{F}')$ *to* $\mathrm{Pop}^c(W, \mathcal{F}, \mathcal{F}')$.

Proof: Again, the difficulty comes in showing that $F_{S \to P}$ is onto. As it says in the main text, given a Popper space $(W, \mathcal{F}, \mathcal{F}', \mu)$, the idea is to construct sets $U_0, U_1, \ldots$ and an LPS $\vec{\mu}$ such that $\mu_\beta(V) = \mu(V \,|\, U_\beta)$, and show that $F_{S \to P}(W, \mathcal{F}, \vec{\mu}) = (W, \mathcal{F}, \mathcal{F}', \mu)$. The construction is somewhat involved.

As a first step, 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: $\inf\{\mu(V_0 \,|\, V_0 \cup V') : V' \in [V]\} > 0$.

Proof: Suppose that $\inf\{\mu(V_0 \,|\, V_0 \cup V') : V' \in [V]\} = 0$. Then there exist sets $V_1, V_2, \ldots$ such that $\mu(V_0 \,|\, V_0 \cup V_n) < 1/n$. 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,

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

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

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

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

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

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

contradicting the choice of α_V. Thus, $\mu(V^* \,|\, V' \cup V^*) = 1$ 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 $\mu(V \,|\, U \cup V) = 0$. 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,

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

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 $\mu(V_j \,|\, \cup_{i=0}^{\infty} U_i) \leq \mu(U_j \,|\, \cup_{i=0}^{\infty} U_i) = 0$. It now follows that using countable additivity that

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

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

Because $\leq$ is well-founded, there is an order-preserving bijection 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 $\mu(U_\beta \,|\, U \cup U_\beta) = 1$ for all $U' \in [U]$. Define an LPS $\vec{\mu} = (\mu_0, \mu_1, \ldots)$ of length α by taking $\mu_\beta(V) = \mu(V \,|\, U_\beta)$. 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 $F_{S\to P}$ 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 3.9: *The map $F_{S\to P}$ is a surjection from* $\mathrm{SLPS}^c(W, \mathcal{F}, \mathcal{F}')$ *onto* $T^c(W, \mathcal{F}, \mathcal{F}')$.

Proof: Suppose that $\mu \in T^c(W, \mathcal{F}, \mathcal{F}')$. I want to construct an SLPS $\vec{\mu} \in SLPS^c(W, \mathcal{F}, \mathcal{F}')$ such that $F_{S\to P}(\vec{\mu}) = \mu$. I first label each element of $\mathcal{F}'$ with a natural number. Intuitively, if $U \in \mathcal{F}'$ is labeled k, then k will be the least index such that $\mu_k(U) > 0$. The labeling is done by induction on k. Each topmost set in the forest (i.e., the root of some tree in the forest) is labeled 0, as are all sets U' such that $\mu(U' \mid U) > 0$, where U is a topmost node. These are all the nodes labeled by 0. Label all the maximal unlabeled sets by 1 (that is, label $U \in \mathcal{F}'$ by 1 if it is not labeled 0, and is not a subset of another unlabeled set); in addition, label a set U' by 1 if $\mu(U' \mid U) > 0$ and U is labeled by 1. Note that every set at depth 0 or 1 in the forest is labeled by either 0 or 1.

Suppose that the labeling process has been completed for labels $0, \ldots, k$ such that the following properties hold, where *label*(U) denotes the label of the event U:

- all sets up to depth k in the forest have been labeled;
- if $label(U) = k'$, $U' \in \mathcal{F}'$, and $\mu(U' \mid U) > 0$, then $label(U') \leq label(U)$.

Label all the maximal unlabeled sets with $k+1$; in addition, if U' is unlabeled and $\mu(U' \mid U) > 0$ for some U such that $label(U) = k+1$, then assign label $k+1$ to U'. Clearly the two properties above continue to hold. This completes the labeling process.

Let $\mathcal{C}_k$ be the set of maximal sets in $\mathcal{F}'$ labeled k. T2 and T3 guarantee that, for all k, the sets in $\mathcal{C}_k$ are disjoint. Let μ'_k be an arbitrary probability on W such that $\mu'_k(U) > 0$ for all $U \in \mathcal{C}_k$ and $\sum_{U \in \mathcal{C}_k} \mu'_k(U) = 1$. Define an LPS $\vec{\mu} = (\mu_0, \mu_1, \ldots)$ as follows (where the length of $\vec{\mu}$ is ω if $\mathcal{C}_k \neq \emptyset$ for all k, and is $k+1$ if k is the largest integer such that $\mathcal{C}_k \neq \emptyset$). For $V \in \mathcal{F}$, let $\mu_j(V) = \sum_{U \in \mathcal{C}_j} \mu(V \mid U)\mu'_j(U)$. I now show that $\vec{\mu}(V \mid U) = \mu(V \mid U)$ for all $V \in \mathcal{F}$ and $U \in \mathcal{F}'$. Suppose that $U \in \mathcal{C}_k$. Then $\mu_j(U) = 0$ for all $j < k$, and $\mu_k(U) > 0$. Thus, $\vec{\mu}(V \mid U) = \mu_k(V \mid U)$. But it is immediate from the definition that $\mu_k(V \mid U) = \mu(V \mid U)$. Thus, $F_{S \rightarrow P}(\vec{\mu}) = \mu$. Moreover, if $U \in \mathcal{F}'$ and $label(U) = k$, let U' be the maximal set containing U such that $label(U') = k$. (The labeling guarantees that such a set exists.) Then $\mu_k(U') = \mu(U' \mid U) > 0$. It follows that $\vec{\mu}(U) > 0$ for all $u \in \mathcal{F}'$. Finally, note that $\vec{\mu}$ is an SLPS (in fact, an LCPS). If $U_k = \cup\mathcal{C}_k - \cup_{k'>k}(\cup\mathcal{C}_{k'})$, then the sets U_k are disjoint, and $\mu_k(U_k) = 1$. ∎

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 $E_\nu(\chi_{U \cap V} - r\chi_U) > E_\nu(\chi_\emptyset)$ iff $r < st(\nu(V \mid U))$. Similarly, $E_{\vec{\mu}}(\chi_{U \cap V} - r\chi_U) > E_{\vec{\mu}}(\chi_\emptyset)$ iff $r < \mu_j(U)$, 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 $\vec{\mu}, \vec{\mu}' \in \mathrm{SLPS}(W, \mathcal{F})$, then $\vec{\mu} \approx \vec{\mu}'$ iff $\vec{\mu} = \vec{\mu}'$.*

Proof: Clearly $\vec{\mu} = \vec{\mu}'$ implies that $\vec{\mu} \approx \vec{\mu}'$. For the converse, suppose that $\vec{\mu} \approx \vec{\mu}'$ for $\vec{\mu}, \vec{\mu}' \in SLPS(W, \mathcal{F})$. If $\vec{\mu} \neq \vec{\mu}'$, let α be the least ordinal such that $\mu_\alpha \neq \mu'_\alpha$, and let U be such that $\mu_\alpha(U) \neq \mu'_\alpha(U)$. Without loss of generality, suppose that $\mu_\alpha(U) > \mu'_\alpha(U)$. Let the sets U_β be such that $\mu_\beta(U_\beta) = 1$ and $\mu_\beta(U_\gamma) = 0$ if $\gamma > \beta$; similarly choose the sets U'_β. Since $\mu_\beta = \mu'_\beta$ for $\beta < \alpha$, it follows that $\mu_\beta(U_\alpha \cup U'_\alpha) = \mu'_\beta(U_\alpha \cup U'_\alpha) = 0$ for $\beta < \alpha$; moreover, $\mu_\alpha(U_\alpha \cup U'_\alpha) = \mu'_\alpha(U_\alpha \cup U'_\alpha) = 1$. Choose r such that $\mu_\alpha(U) > r > \mu'_\alpha(U)$. Let X be the random variable $\chi_U - r\chi_{U_\alpha \cup U'_\alpha}$ and let $Y = \chi_\emptyset$. Then $E_{\vec{\mu}}(X) > E_{\vec{\mu}}(Y)$, while $E_{\vec{\mu}'}(X) < E_{\vec{\mu}'}(Y)$, so $\vec{\mu} \not\approx \vec{\mu}'$. ∎

Proposition 4.4: *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 $\vec{\mu}' = (\mu_{k_0}, \ldots, \mu_{k_h})$ 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 4.5: *If W is finite, then $F_{L\to N}$ is a bijection from* LPS$(W, \mathcal{F})/\!\approx$ *to* 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 provide a sufficient condition for an NPS to be equivalent an LPS in a finite space.

Lemma A.7: *Suppose that $\vec{\mu} = (\mu_0, \ldots, \mu_k)$, and $\epsilon_0, \ldots, \epsilon_k$ 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 $\vec{\mu} \approx \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$.*[11]

Proof: Suppose that there exist $\epsilon, \ldots, \epsilon_k$ as in the statement of the lemma and $\nu = \epsilon_0\mu_0 + \cdots + \epsilon_k\mu_k$. I want to show that $\vec{\mu} \approx \nu$.

If $E_{\vec{\mu}}(X) < E_{\vec{\mu}}(Y)$, then there exists some $j \le k$ 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 $\epsilon_j(E_{\mu_j}(Y) - E_{\mu_j}(X)) > \sum_{i=j+1}^k \epsilon_i(E_{\mu_i}(X) - E_{\mu_i}(Y))$. Since $\epsilon_{j'+1} \le \epsilon_{j'}$ for $j' \ge j$ (this follows from the fact that $st(\epsilon_{j'+1}/\epsilon j') = 0$), it follows that $\sum_{i=j+1}^k \epsilon_i(E_{\mu_i}(X) - E_{\mu_i}(Y)) \le \epsilon_{j+1}\sum_{i=j+1}^k |E_{\mu_i}(X) - E_{\mu_i}(Y)|$. Thus, it suffices to show that $\epsilon_{j+1}\sum_{i=j+1}^k |E_{\mu_i}(X) -$

[11] Although I do not need this fact here, it is easy to see that if W is finite and $\vec{\mu} = (\mu_0, \ldots, \mu_k)$ is an SLPS in $LPS(W, \mathcal{F})$, then the converse of Lemma A.7 holds as well: if $\nu \approx \vec{\mu}$, then $\nu = \epsilon_0\mu_0 + \cdots \epsilon_k\mu_k$ for some $\epsilon_0, \ldots, \mu_k$ are such that $st(\epsilon_{i+1}/\epsilon_i) = 0$ for $i = 1, \ldots, k-1$ and $\sum_{i=0}^k \epsilon_i = 1$. (I conjecture this fact is true in general, not just if $\vec{\mu}$ is an SLPS, but I have not checked this.

$E_{\mu_i}(Y)| < \epsilon_j(E_{\mu_j}(Y) - E_{\mu_j}(X))$. This is trivially the case if $E_{\mu_i}(X) = E_{\mu_i}(Y)$ for all i such that $j+1 \leq i \leq k$. Thus, assume without loss of generality that $\sum_{i=j+1}^{k} |E_{\mu_i}(X) - E_{\mu_i}(Y)| > 0$. In this case, it suffices to show that $\epsilon_{j+1}/\epsilon_j < (E_{\mu_j}(Y) - E_{\mu_j}(X))/\sum_{i=j+1}^{k} |E_{\mu_i}(X) - E_{\mu_i}(Y)|$. Since the right-hand side of the inequality 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 $E_\nu(X) = \sum_{i=0}^{k} \epsilon_i E_{\mu_i}(X)$ and $E_\nu(Y) = \sum_{i=0}^{k} \epsilon_i E_{\mu_i}(Y)$, 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 that $\vec{\mu} \approx \nu$. ∎

It remains to show that, given an NPS $(W, \mathcal{F}, \nu)$, there is an equivalence class $[\vec{\mu}]$ such that $F_{L \to N}([\vec{\mu}]) = [\nu]$. As I said in the main text, the goal now 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 $Basic(\mathcal{F}) = \{U_1, \ldots, U_k\}$ 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 $0 \leq m < k$. 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 $\vec{b}_m = st\left(\vec{b}'_m\right)$ 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 4.5, 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 4.6: *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 4.4, using the observation that a probability measure over $(W, \mathcal{F})$ can be identified with an element of $I\!R^{|\mathcal{F}|}$; 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 4.10, 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

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

or

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

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 bijection from* $\mathrm{NPS}(W, \mathcal{F})/\simeq$ *to* $\mathrm{Pop}(W, \mathcal{F})$ *and from* $\mathrm{NPS}^c(W, \mathcal{F})/\simeq$ *to* $\mathrm{Pop}^c(W, \mathcal{F})$.

Proof: As I said in the main text, the proof that $F_{N\to P}$ is an injection is straightforward, and to prove that it is a surjection in the countably additive case, it suffices to show that $F_{N\to P}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$, where $\nu \approx \vec{\mu}'$ and $\vec{\mu}'$ is the countably additive SLPS such that $F_{S\to P}((W, \mathcal{F}, \vec{\mu}')) = (W, \mathcal{F}, \mathcal{F}', \mu)$. I now do this.

Suppose that $F_{N\to P}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}'_1, \mu_1)$. First I show that $\nu(U) = 0$ iff $\vec{\mu}'(U) = \vec{0}$. Let $X = \chi_U$ and $Y = \chi_\emptyset$. Note that $\nu(U) = 0$ iff $E_\nu(X) = E_\nu(Y)$ iff $E_{\vec{\mu}'}(X) = E_{\vec{\mu}'}(Y)$ iff $\vec{\mu}'(U) = \vec{0}$. Thus, $\mathcal{F}'_1 = \{U : \nu(U) \neq 0\} = \{U : \vec{\mu}'(U) \neq \vec{0}\} = \mathcal{F}'$.

Now suppose by way of contradiction that $\mu \neq \mu_1$. Thus, there must exist some $V \in \mathcal{F}$, $U \in \mathcal{F}'$ such that $\mu(V \mid U) \neq \mu_1(V \mid U)$. Let β be the smallest ordinal such that $\mu'_\beta(U) \neq 0$. It follows that $\mu'_\beta(V \mid U) \neq st(\nu(V \mid U))$. We can assume without loss of generality that $\mu'_\beta(V \mid U) > st(\nu(V \mid U))$. Choose a real number r such that $\mu'_\beta(V \mid U) > r > st(\nu(V \mid U))$. Then $E_{\vec{\mu}'}(\chi_{V\cap U}) > E_{\vec{\mu}'}(r\chi_U)$ but $E_\nu(\chi_{V\cap U}) < E_\nu(r\chi_U)$. This contradicts the assumption that $\vec{\mu}' \approx \nu$. It follows that $F_{N\to P}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$, as desired.

It remains to show that if $(W, \mathcal{F}, \mathcal{F}', \mu) \in Pop(W, \mathcal{F}) - Pop^c(W, \mathcal{F})$, then there is some $(W, \mathcal{F}, \nu) \in NPS(W, \mathcal{F})$ such that $F_{N\to P}(W, \mathcal{F}\nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$. My proof in this case follows closely the lines of an analogous result proved by McGee [1994]. 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 relies on only the fact that the algebra is finite, so it applies without change here.) It now follows from Theorem 4.5 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 $\mathcal{G} = \{G \subseteq J : G \supseteq G_{\mathcal{A}} \text{ for some } \mathcal{A} \subseteq \mathcal{F}\}$. 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 $\nu_1 \approx \nu_2$ then $\nu_1 \simeq \nu_2$.*

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. ∎

Theorem 6.5: *There exists an NPS ν whose range is an elementary extension of the reals such that $\vec{\mu} \approx \nu$ and $X_1, \ldots, X_n$ are independent with respect to ν iff there exists a sequence $\vec{r}^j$, $j = 1, 2, \ldots$ of vectors in $(0,1)^k$ such that $\vec{r}^j \to (0, \ldots, 0)$ as $j \to \infty$, and $X_1, \ldots, X_n$ are independent with respect to $\vec{\mu} \,\Box\, \vec{r}^j$ for $j = 1, 2, 3, \ldots$.*

Proof: Suppose that there exists an NPS ν whose range is an elementary extension of the reals, $\vec{\mu} \approx \nu$, and $X_1, \ldots, X_n$ are independent with respect to ν. Using arguments similar in spirit to those the arguments of BBD [1991b, Proposition 2], it follows that there exist positive infinitesimals $\epsilon_1, \ldots, \epsilon_k$ such that $\vec{\mu} \,\Box\, (\epsilon_1, \ldots, \epsilon_k) = \nu$. It is not hard to show that there exists a finite set of real-valued polynomials $p_1, \ldots, p_N$ such that $p_j(\epsilon_1, \ldots, \epsilon_k) = 0$ for $j = 1, \ldots, N$, and if $\vec{r}$ is a vector of positive reals such that $p_j(\vec{r}) = 0$ for $j = 1, \ldots, N$, then $X_1, \ldots, X_n$ are independent with respect to $\vec{\mu} \,\Box\, \vec{r}$. Thus, for all natural numbers $m \geq 1$, the range of ν satisfies the first-order property

$$\exists x_1 \ldots \exists x_k (p_1(x_1, \ldots, x_k) = 0 \wedge \ldots \wedge p_N(x_1, \ldots, x_k) = 0 \wedge 0 < x_1 < 1/m \wedge \ldots \wedge 0 < x_k < 1/m).$$

Since the range of ν is an elementary extension of the reals, this first-order property holds of the reals as well. Thus, there exists a sequence $\vec{r}^j$ of vectors of positive reals converging to $\vec{0}$ such that $p_j(\vec{r}^j) = 0$ for $j = 1, \ldots, N$.

The converse follows by a straightforward application of compactness in first-order logic [Enderton 1972]. Suppose that there exists a sequence $\vec{r}^j$, $j = 1, 2, \ldots$ of vectors in $(0,1)^k$ such that $\vec{r}^j \to (0, \ldots, 0)$ as $j \to \infty$, and $X_1, \ldots, X_n$ are independent with respect to $\vec{\mu} \,\Box\, \vec{r}^j$ for $j = 1, 2, 3, \ldots$. We now apply the compactness theorem. As I mentioned in the proof of Proposition 4.6, 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 the function symbols $+$ and $\times$, the binary relation $\leq$, a constant symbol $\mathbf{r}$ for each real number r, a unary predicate N (representing the natural numbers), and constant symbols p_U for each set $U \in \mathcal{F}$. Intuitively, p_U represents $\nu(U)$. Consider the following (uncountable) collection of formulas:

(a) All first-order formulas in this language true of the reals. (This includes, for example, a formula such as $\forall x \forall y(x+y=y+x)$, which says that addition is commutative, as well as formulas such as $\mathbf{2}+\mathbf{3}=\mathbf{5}$ and $\sqrt{\mathbf{2}} \times \sqrt{\mathbf{3}}=\sqrt{\mathbf{6}}$.)

(b) Formulas $p_U > 0$ for $U \in \mathcal{F}'$ and $p_U = 0$ for $U \in \mathcal{F} - \mathcal{F}'$.

(c) Formulas $p_U + p_V = p_{U \cup V}$ if $U \cap V = \emptyset$.

(d) The formula $p_W = 1$.

(e) Formulas of the form $p_{X_1=x_1} \times \cdots \times p_{X_n=x_n} = p_{X_1=x_1 \cap \ldots \cap X_n=x_n}$, for all values $x_i \in \mathcal{V}(X_i)$, $i=1,\ldots,n$; these formulas say that $X_1,\ldots,X_n$ are independent with respect to ν.

(f) For every pair of Y, Y' of random variables such that $E_{\vec{\mu}}(Y) \geq E_{\vec{\mu}}(Y')$, a formula that says $E_\nu(Y) \geq E_\nu(Y')$, where $E_\nu(Y)$ and $E_\nu(Y')$ are expressed using the constant symbols p_U (where the events U are those of the form $Y=y$ and $Y'=y'$). Note that this formula is finite, since X and Y are assumed to have finite range. The formula would not be expressible in first-order logic if X or Y had infinite range.

It is not hard to show that every finite subset of these formulas is satisfiable. Indeed, given a finite subset of formulas, there must exist some m such that taking $p_U = \vec{\mu} \,\Box\, \vec{r}^m(U)$ will work (and interpreting $\mathbf{r}$ as the real number r, of course). The only nonobvious part is showing that we can deal with the formulas in part (f); that we can do so follows from the proof of Proposition 1 in [Blume, Brandenburger, and Dekel 1991b], which shows that $E_{\vec{\mu}}(Y') > E_{\vec{\mu}}(Y)$ iff there exists some M such that $E_{\vec{\mu} \,\Box\, \vec{r}^m}(Y') > E_{\vec{\mu} \,\Box\, \vec{r}^m}(Y)$ for all m, then $E_{\vec{\mu}}(Y') > E_{\vec{\mu}}(Y)$.

Since every finite set of formulas is satisfiable, by compactness, the infinite set is satisfiable. Let $\nu(U)$ be the interpretation of p_U in a model satisfying these formulas. Then it is easy to check that ν is an elementary extension of the reals, $\nu \approx \vec{\mu}$, and that $X_1,\ldots,X_n$ are independent with respect to ν. ∎

Theorem 6.6: *$X_1,\ldots,X_n$ are strongly independent with respect to the Popper space $(W,\mathcal{F},\mathcal{F}',\mu)$ iff there exists an NPS $(W,\mathcal{F},\nu)$ such that $F_{N\to P}(W,\mathcal{F},\nu) = (W,\mathcal{F},\mathcal{F}',\mu)$ and $X_1,\ldots,X_n$ are independent with respect to $(W,\mathcal{F},\nu)$.*

Proof: It easily follows from Kohlberg and Reny's [1997, Theorem 2.10] characterization of strong independence that if $X_1,\ldots,X_n$ are independent with respect to the NPS $(W,\mathcal{F},\nu)$ then $X_1,\ldots,X_n$ are strongly independent with respect to $F_{N\to P}(W,\mathcal{F},\nu)$.

The converse follows using compactness, much as in the proof of Theorem 6.5. Suppose that $(W,\mathcal{F},\mathcal{F}',\mu)$ is a Popper space and $\mu_j \to \mu$ are as required for $X_1,\ldots,X_n$ to be strongly independent with respect to μ. Consider the same language as in the proof of Theorem 6.5, and essentially the same collection of formulas, except that the formulas of part (f) are replaced by (f′) Formulas of the form $(\mathbf{r} - \frac{\mathbf{1}}{\mathbf{n}})\mathbf{p_V} \le \mathbf{p_{U\cap V}} \le (\mathbf{r} + \frac{\mathbf{1}}{\mathbf{n}})\mathbf{p_V}$ for all U, V, $\mathbf{r}$, and $\mathbf{n} > \mathbf{0}$ such that $\mu(U \mid V) = r$.

Again, it is easy to see that every finite subset of these formulas is satisfiable. Indeed, given a finite subset of formulas, there must exist some m such that taking $p_U = \mu_m(U)$ satisfies all the formulas (and interpreting $\mathbf{r}$ as the real number r, of course). By compactness, the infinite set is satisfiable. Let $\nu(U)$ be the interpretation of p_U in a model satisfying these formulas. Then it is easy to check that $F_{L\to N}(W, \mathcal{F}, \nu) = (W, \mathcal{F}, \mathcal{F}', \mu)$, and that $X_1, \ldots, X_n$ are independent with respect to ν. ∎

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.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *Annals of Mathematical Statistics 34*, 199–205.

Asheim, G. B. (2006). *The Consistent Preferences Approach to Deductive Reasoning in Games*. Berlin/Heidelberg/New York: Springer.

Asheim, G. B. and A. Perea (2005). Sequential and quasi-perfect rationalizability in extensive games. *Games and Economic Behavior 53*, 15–42.

Asheim, G. B. and Y. Søvik (2005). Preference-based belief operators. *Mathematical Social Sciences 50*, 61–82.

Battigalli, P. (1996). Strategic independence and perfect Bayesian reasoning. *Journal of Economic Theory 70*, 201–234.

Battigalli, P. and M. Siniscalchi (1999). Interactive beliefs, epistemic independence, and strong rationalizability. *Research in Economics 53*, 247–273.

Battigalli, P. and M. Siniscalchi (2002). Strong beliefs and forward-induction reasoning. *Journal of Economic Theory 106*, 356–391.

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., A. Friedenberg, and J. Keisler (2008). Admissibility in games. *Econometrica 76*(2), 307–352.

Coletti, G. and R. Scozzafava (2002). *Probabilistic Logic in a Coherent Setting*. Dordrecht, Netherlands: Kluwer.

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

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.

Fudenberg, D. and J. Tirole (1991). Perfect Bayesian reasoning. *Journal of Economic Theory 53*, 236–260.

Halpern, J. Y. (2001). Lexicographic probability, conditional probability, and nonstandard probability. In *Theoretical Aspects of Rationality and Knowledge: Proc. Eighth Conference (TARK 2001)*, pp. 17–30. San Francisco: Morgan Kaufmann. A full version of the paper can be found at http://arxiv.org/abs/cs/0306106.

Halpern, J. Y. (2002). Characterizing the common prior assumption. *Journal of Economic Theory 106*(2), 316–355.

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*, pp. 25–49. Dordrecht, Netherlands: Kluwer.

Hammond, P. J. (1999). Consequentialism, non-Archimedean probabilities, and lexicographic expected utility. In C. Bicchieri, R. Jeffrey, and B. Skyrms (Eds.), *The Logic of Strategy*. Oxford, U.K.: Oxford University Press.

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.

Kohlberg, E. and P. J. Reny (1997). Independence of relative probability spaces and consistent assessments in game trees. *Journal of Economic Theory 75*, 280–313.

Krauss, P. H. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae 19*(3–4), 229–241.

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. (1934). *Logik der Forschung*. Vienna: Julius Springer Verlag.

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.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*, 25–55.

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

Stalnaker, R. (1998). Belief revision in games: forward and backward induction. *Mathematical Social Sciences 36*, 31–56.

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.