QED

Queen's Economics Department Working Paper No. 874

Logics for Nonomniscient Agents: An Axiomatic Approach

Barton L. Lipman

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-1993

Discussion Paper #874

Logics for Nonomniscient Agents: An Axiomatic Approach

by

Barton L. Lipman
Queen's University

January 1993

Logics for Nonomniscient Agents: An Axiomatic Approach

by

Barton L. Lipman

Department of Economics
Queen's University
Kingston, Ontario K7L 3N6

email: lipmanb@qucdn.queensu.ca

Queen's University Discussion Paper #874

First Draft

January 1993

I am grateful for comments on this and a related paper from seminar audiences at the Second Workshop on Knowledge, Belief, and Strategic Interaction, Summer in Tel Aviv 1992, Queen's, Cornell, Waterloo, York, Victoria, and University of British Columbia. I also thank Jim Bergin, Itzhak Gilboa, Debra Holt, Michele Piccione, and Aldo Rustichini. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. Of course, any errors are my own responsibility.

Abstract.

It has long been recognized that solving the logical omniscience problem requires using some kind of nonstandard possible worlds. While many such logics have been proposed, none has an obvious claim as the "right" logic to use to describe the reasoning of agents who are not logically omniscient. I show how to derive such nonstandard worlds as part of a representation of an agent's preferences. In this sense, the agent's logic is given the same basis as a utility function or subjective probability. As an illustration, I give conditions on preferences which imply that the agent's logic is a version of the logic of inconsistency proposed by Rescher and Brandom [1979].

Key Words: Bounded rationality, decision theory, nonstandard logics.

JEL Classification Numbers: C7, D81.

I. Introduction.

It has long been known that the standard possible worlds approach to representing knowledge and beliefs has one very important implication, dubbed by Hintikka [1975] the problem of *logical omniscience*. The possible worlds approach says that an agent knows that p is true if and only if p is true in every world the agent conceives of as possible. Suppose the agent learns that p is true where $p \to q$ is a tautology. If every world the agent conceives of as possible is logically consistent, then $p \to q$ must be true in every such world. Hence in any such world, if p is true, q is true as well. Therefore, an agent who learns that p *must* recognize that q is true. In this sense, the agent knows every logical implication of his knowledge. While this is a very attractive property for the study of ideal reasoners, it is unpalatable as an assumption about real people.

I believe that game theorists should also be very interested in relaxing the logical omniscience assumption. Many of the examples we find to be "paradoxical," such as the centipede game (Rosenthal [1981], Reny [1986], Binmore [1987]) or van Damme's [1989] dollar–burning example, rely on a complex deduction from a simple and plausible set of hypotheses. We believe that the hypotheses may well be known to a real agent, but we are reluctant to believe that a real agent would reach the conclusion. It is precisely the assumption of logical omniscience which makes this view difficult to formalize in standard models.

Fortunately, there is a simple — even obvious — solution to the problem. If some of the worlds the agent conceives of as possible are *not* logically consistent, then the chain of reasoning above is broken. If the agent conceives of a world in which $p \to q$ is true, p is true, but q is false, then learning p does not lead the agent to recognize that q is true, even if he already knows that $p \to q$. Such worlds go by a variety of names, including *nonstandard possible worlds* (Rescher and Brandom [1979]) or *impossible possible worlds* (Hintikka [1975], Rantala [1975]). I will use these terms interchangably.

The difficulty with this solution, unfortunately, is also quite obvious: what should we assume about the impossible possible worlds? Put differently, exactly which nonstandard logic should we use to describe the reasoning of real agents? It is quite clear what "perfect reasoning" entails; it is not at all obvious how to give a precise formulation of "imperfect reasoning."

In this paper, I propose an approach to this problem which I study further in Lipman [1992a]. The idea is to derive the nonstandard worlds the agent conceives of as possible — and hence to derive the agent's "logic" — by analyzing his preferences. In a sense, then, the agent's logic is derived as a representation of preferences in the same way a utility function or subjective probabilities would be derived. Intuitively, if the agent's reasoning does indeed affect his choices, this effect must be observable in some fashion. The natural place to look for this effect is the agent's preferences, or, more specifically, the way the agent's preferences vary with his information.

The simplest way to see this clearly is to suppose that p and q are logically equivalent propositions. Suppose, though, that the agent does not respond to these pieces of information in the same way. That is, his preferences if he is told that p is true (and is told nothing about q directly) differ from the preferences he has if he is told that q is true (and is told nothing about p directly). Then we can infer that the agent does not recognize the fact that p and q are logically equivalent. Hence there must be at least one impossible possible world for this agent in which one of the two propositions is true and the other is not.

The rest of this paper is organized as follows. In Section II, I give the basic framework for introducing impossible possible worlds. I also prove a simple theorem which shows that introducing impossible possible worlds allows us to "rationalize" virtually any preferences. I see this as both a negative result and a positive one. On the one hand, it is certainly undeniable that a model which rationalizes everything explains nothing. On the other hand, this result means that this kind of approach can nest virtually any sort of preferences we might wish to discuss. In this sense, it gives us a sufficiently broad "language" for viewing the problem.

In Section III, I illustrate the use of the framework by giving conditions on the agent's preferences which are necessary and sufficient for attributing to him a version of the logic of inconsistency proposed by Rescher and Brandom [1979]. Section IV offers some concluding remarks. All proofs are contained in the Appendix.

Related Literature. There are several papers which bear very strong connections to this work. Gilboa and Schmeidler [1992] show that Choquet expected utility — that is, expected utility with respect to a nonadditive probability measure — is equivalent to expected utility on an enlarged state space. Their enlargement of the state space can be seen as introducing impossible possible worlds. Consequently, some of my results are generalizations of theirs, as I will point out later. Second, it is well–known that Shafer's [1976] belief functions — introduced as an alternative to probability for representing uncertainty — can be derived from additive functions on a larger state set. Again, this can be seen as a particular way to introduce impossible possible worlds. Third, as noted earlier, many approaches to constructing impossible possible worlds or nonstandard logics have been proposed in the philosophy and artifical intelligence literature. See, for example, Rantala [1975], Fagin and Halpern [1985], or Fagin, Halpern, and Vardi [1990]. While I focus on the logic proposed by Rescher and Brandom, clearly, a similar analysis could be carried out for a variety of other nonstandard logics. Finally, Morris [1992] provides an axiomatic derivation of nonpartitional information structures which has some similarities to my work. In particular, he also uses the way preferences vary with information to derive statements about the agent's reasoning. For more details on the relationship between Morris' work and my own, see Lipman [1992a, 1993].

II. Framework for Analysis.

To model the way preferences vary with information, I need a model of information which does not presume logical omniscience. Hence I begin with propositions as abstract variables, rather than sets of possible worlds. So let P denote the set of relevant propositions. I assume that P is nonempty and is closed under $\neg$, $\vee$, $\wedge$, and $\rightarrow$ (which are "not," "or," "and," and "implies" respectively). That is, if $p \in P$, then $\neg p \in P$ and if $p, q \in P$, then $p \vee q \in P$, $p \wedge q \in P$, and $p \rightarrow q \in P$.[1]

The next ingredient I require is a notion of what "correct" logical deduction is. This is modeled as follows. A state of the world will be a collection of propositions which constitutes a complete and logically consistent description of how the world might be, listing all the facts that are true in some possible complete description. More formally, $s \subseteq 2^P$ is a *state of the world* or a *possible world* if

$$p \in s \iff \neg p \notin s$$

$$p \vee q \in s \iff p \in s \text{ or } q \in s$$

$$p \wedge q \in s \iff p \in s \text{ and } q \in s$$

$$p \rightarrow q \in s \iff \neg p \in s \text{ or } q \in s$$

Let S denote the set of all possible worlds. For simplicity, I assume S is finite.

For any $p \in P$, let $S(p) = \{s \in S \mid p \in s\}$. That is, $S(p)$ is the collection of states of the world in which p is true. If $S(p) = S$, I will say that p is *tautologically true* or a *tautology*. If $S(p) = \emptyset$, p is *tautologically*

---

1 Note that I do not rule out the possibility of quantifiers in the language, though I do not analyze them. Rescher and Brandom [1979] give some intriguing examples with quantifiers, suggesting that further analysis may be of interest.

false. By construction, $S(p \vee \neg p) = S$ for all p, while $S(p \wedge \neg p) = \emptyset$ for all p. For convenience, I assume that there is a special proposition $\top \in P$ such that $\top$ is tautologically true and a special proposition $\perp$ such that $S(\perp) = \emptyset$. Let $\bar{P} = \{p \in P \mid S(p) \neq \emptyset\}$.

Set–Theoretic Notational Conventions. For any sets A and B, I will let A^B denote the set of all functions $f : B \to A$. Similarly, 2^B will denote the set of all subsets of B. Also, $\#B$ will denote the cardinality of B. If B is a collection of sets, then $\cap B$ is the intersection of all the sets in B and $\cup B$ is the union of all the sets in B. Finally, I will use $\subseteq$ to denote weak containment (*i.e.*,, subset or equal to) and reserve $\subset$ for *strict* containment.

Let X be the set of *consequences*. As in Savage [1954], the interpretation of a consequence is that it is as complete a specification of the outcome of a choice as is necessary to describe an agent's evaluation of that outcome. For simplicity, I will take $X = \mathbf{R}$. Let $F = X^S$ denote the set of *acts*. That is, an act or action is viewed in terms of the relationship it creates between external events (which propositions hold) and consequences.

For each $p \in \bar{P}$, we have a preference relation on F, $\succ_p$, to be interpreted as the agent's preferences given that he learns that p is true. That is, $f \succ_p g$ is interpreted as saying that the agent would strictly prefer act f to act g if he learns that proposition p is true. Let $\{\succ_p\}$ denote the collection of these preference orderings and let $\succ \; = \; \succ_\top$. I emphasize that I make no assumption about the agent's self–awareness. I assume that we, as modellers, know how the agent would respond to each possible piece of information, not that the agent himself knows this *ex ante*. Note also that I do not ask for information about the agent's preferences in response to nonsensical pieces of information such as $p \wedge \neg p$.

A natural way to try to represent these preferences would be with expected utility. To state this more precisely, I will say that a proposition $p \in P$ is *null* if either $p \notin \bar{P}$ or $f \sim_p g$ for all $f, g \in F$.

Definition. $\{\succ_p\}$ *is EUR* — expected utility representable — *if there is a function* $u : X \to \mathbf{R}$ *and a probability measure* μ *on* S *such that for all nonnull* p, $\mu(S(p)) > 0$ *and*

$$f \succ_p g \iff \mathrm{E}_\mu[u(f(s)) \mid s \in S(p)] > \mathrm{E}_\mu[u(g(s)) \mid s \in S(p)],$$

where $\mathrm{E}_\mu[\cdot \mid s \in S(p)]$ *denotes the expectation with respect to the measure* μ *conditional on the event* $s \in S(p)$.

It is straightforward to restate the Savage [1954] axioms in this framework to give sufficient conditions for such a representation.[2]

There is one necessary condition for an expected utility representation which is implicit in the usual framework and so is not normally discussed. Say that p and q are *logically equivalent* if $S(p) = S(q)$. It is not hard to see that if $\{\succ_p\}$ is EUR, then for any logically equivalent p and q, we must have $\succ_p = \succ_q$. To see this, simply note that the effect of conditioning on p is the same as the effect of conditioning on q. That is, the expectation conditional on $s \in S(p)$ must be the same as the expectation conditional on $s \in S(q)$ if $S(p) = S(q)$.

Definition. $\{\succ_p\}$ *satisfies* revealed logical omniscience *(or RLO) if* $S(p) = S(q)$ *implies* $\succ_p = \succ_q$.

If $\{\succ_p\}$ can be represented by expected utility, then it must satisfy RLO.

Several comments on the definition of RLO are in order. First, note that the agent's preferences may satisfy RLO without the agent being aware of all logical implications of his knowledge. In this sense, he may not be "truly" logically omniscient. However, he is *effectively* logically omniscient.

---

[2] The finiteness of the state space does complicate matters. See Gul [1992] and Chew and Karni [1992].

Second, note that the definition of RLO refers only to logical equivalence, not logical implication. As we will see, there is not a behaviorial distinction between this notion of logical omniscience and the apparently stronger idea that the agent recognizes all logical implications. Third, as we will see, my terminology may be slightly misleading. While I have used "logical omniscience" in the name of this condition, it does *not* imply that the agent reasons perfectly.

Intuitively, the reason the preferences of real agents are unlikely to satisfy RLO is that they do not recognize logical equivalence. That is, the problem is not that the agent knows that $S(p) = S(q)$ but wishes to behave differently when learning that p is true than when learning q. Instead, it is that the agent simply doesn't realize that $S(p) = S(q)$. This suggests the following approach.

Definition. *$\{\succ_p\}$ is XEUR* — extended expected utility representable — *if there exists $S^* \subseteq 2^P$, $h : F \to X^{S^*}$, $u : X \to \mathbf{R}$, and a probability measure μ on S^* such that*

$$S \subseteq S^*; \quad \top \in s^*, \ \bot \notin s^* \quad \forall s^* \in S^*,$$

$$h(f)|S = f, \quad \forall f \in F,$$

and for all nonnull p, $\mu(S(p)) > 0$ and

$$f \succ_p g \iff \mathrm{E}_\mu[u(h(f)(s)) \mid s \in S^*(p)] > \mathrm{E}_\mu[u(h(g)(s)) \mid s \in S^*(p)]$$

where $\mathrm{E}_\mu[\cdot \mid s \in S^*(p)]$ is the conditional expectation with respect to μ given the event $s \in S^*(p)$ and

$$S^*(p) = \{s \in S^* \mid p \in s\}.$$

In other words, $\{\succ_p\}$ is XEUR if we can extend the state set from S to S^* — *i.e.*, introduce impossible possible worlds — and extend all lotteries to the new state set (via the function h) in such a way that the preferences are represented by expected utility on the larger state set. (In the definition, $h(f)|S$ is the restriction of the function $h(f)$ to S, so the definition requires that $h(f)(s) = f(s)$ for all $s \in S$.) The requirement that $\top \in s^*$ for all s^* simply guarantees that there is still a meaningful notion of preferences without any information. The requirement that $\bot \notin s^*$ for any s^* is primarily a matter of notational convenience.

Remark 1. XEUR requires that $\mu(S(p)) > 0$ if p is nonnull. This is stronger than the perhaps more natural requirement that $\mu(S^*(p)) > 0$ (note that $S(p) \subseteq S^*(p)$). However, without this requirement, XEUR becomes completely trivial since one can choose a μ such that $\mu(S) = 0$. In this case, only the extension is relevant to the preferences.

Theorem 1 below essentially says that almost any "reasonable" collection of preferences is XEUR. As the proof shows, this result is due to the fact that the extension can be chosen in such a way that the evaluation of lotteries on $S^* \setminus S$ determines the evaluation overall. Thus, with no restrictions on how the lotteries are extended, virtually any preference can be represented.

Definition. $\{\succ_p\}$ *is* representable *if for all nonnull p, there exists $u_p : F \rightarrow$* **R** *such that*

$$f \succ_p g \iff u_p(f) > u_p(g).$$

Clearly, representability is necessary for $\{\succ_p\}$ to be XEUR — without it, preferences conditional on some nonnull p are not representable by a utility function at all, much less one with the particular structure XEUR requires. Necessary and sufficient conditions for representability are well known so I will omit discussion of them. It is sufficient to note that representability is vastly weaker than the axioms needed for EUR.

Theorem 1 says that representability, together with one simplifying assumption, is sufficient for XEUR. Hence representability is essentially necessary and sufficient for XEUR. The auxiliary assumption I will make is:

(A1) There are finitely many distinct preference relations in the collection $\{\succ_p\}$.

Recall that S is finite. (A1) and the finiteness of S can certainly be relaxed but these assumptions enable me to avoid constructing a σ–algebra.

Theorem 1. *If representability and (A1) hold, then $\{\succ_p\}$ is XEUR.*

The theorem is proved by constructing impossible possible worlds and a way of extending the acts given a particular collection of preferences. To understand the idea, say that p and q are ***strongly equivalent*** if they are logically equivalent and $\succ_p = \succ_q$. When two propositions are strongly equivalent, I include either both or neither of them in each impossible possible world. That is, strongly equivalent propositions are ones which the agent recognizes as equivalent. If two propositions are logically equivalent but not strongly equivalent, then impossible possible worlds must be introduced in which one but not the other is true. Of course, this must be done in such a way as to represent the particular preferences in question.

III. An Axiomatic Derivation of the Logic of Inconsistency.

In this section, I give conditions on the agent's preferences which are necessary and sufficient to represent him by assuming that his deductions are based on a version of Rescher and Brandom's [1979] logic of inconsistency. I proceed as follows. First, I show how to define two deduction relations,[3] one which describes the "true" logic (and is entirely standard) and the other

[3] I do not use the more common term "consequence relation" to avoid confusion with the consequence set X.

of which describes the agent's logic. That is, the latter can be interpreted as our representation of the logic which underlies the agent's deductions. Next, I will present a slight generalization of the logic of inconsistency proposed by Rescher and Brandom and derive some of its properties. Finally, I use this information to relate statements about the agent's preferences to statements about his logic. Throughout this section, S^* will denote a subset of 2^P such that $S \subseteq S^*$, $S^*(\top) = S^*$, and $S^*(\bot) = \emptyset$.

First, I will define the "true" logic in the form of a deduction relation $\vdash$. My approach will differ from the usual one in two ways, neither of which is important in the standard logic, but both of which are important in the agent's logic. Normally, one writes $p_1, \ldots, p_n \vdash q$ if from the premises $p_1, \ldots, p_n$, one can deduce q. The first important issue concerns what we mean by "the premises $p_1, \ldots, p_n$." I will interpret this to mean the collection of propositions rather than to mean the conjunction $p_1 \wedge \ldots \wedge p_n$. The agent may well recognize that if each of the propositions $p_1, \ldots, p_n$ is true, then the conjunction must be true. However, we wish to study agents who may fail to recognize such implications. Consequently, I will not assume that the agent makes the translation from the collection of propositions to their conjunction.

The second important question concerns the deduction. In particular, if the deduction is a single proposition, how can we represent the notion that the agent recognizes that either q_1 or q_2 must be true but does not know which? Normally, we would say that this is the same as his deducing $q_1 \vee q_2$. But again, treating these statements as equivalent presumes that the agent uses the $\vee$ operator correctly in all his deductions, a treatment not obviously appropriate here. For this reason, as in Shoesmith and Smiley [1978], I will allow the conclusion to be a collection of propositions. The interpretation, then, is that $p_1, \ldots, p_n \vdash q_1, \ldots, q_m$ holds if from the premises that each of $p_1, \ldots, p_n$ is true, the agent recognizes that at least one of $q_1, \ldots, q_m$ must be true. Formally,

$$p_1, \ldots, p_n \vdash q_1, \ldots, q_m \iff \bigcap_{i=1}^{n} S(p_i) \subseteq \bigcup_{i=1}^{m} S(q_i).$$

Of course, this definition implies that

$$(1) \qquad p_1, \ldots, p_n \vdash q_1, \ldots, q_m \iff p_1 \wedge \ldots \wedge p_n \vdash q_1 \vee \ldots \vee q_m.$$

Finally, given S^*, I will define $\vdash_*$ to be the logic it implies. That is,

$$p_1, \ldots, p_n \vdash_* q_1, \ldots, q_m \iff \bigcap_{i=1}^{n} S^*(p_i) \subseteq \bigcup_{i=1}^{m} S^*(q_i).$$

In general, $\vdash_*$ will *not* satisfy the analogous condition to (1).

The logic proposed by Rescher and Brandom is based on constructing the new worlds in $S^* \setminus S$ from the old ones in S in a particularly simple way. Since the elements of S are sets, a natural construction procedure to consider is unioning or intersecting these sets to create new states. Naturally, one might also wish to consider unions of intersections of states, etc. The simplest way to allow such possibilities is the following. Recall that for any collection of sets B, $\cap B$ denotes the intersection of the sets in B and $\cup B$ denotes the union. Given a collection of subsets of P, say $\mathcal{P}$, let

$$I(\mathcal{P}) = \{s^* \in 2^P \mid s^* = \cap B, \quad \text{some } B \subseteq \mathcal{P}\}$$

and

$$U(\mathcal{P}) = \{s^* \in 2^P \mid s^* = \cup B, \quad \text{some } B \subseteq \mathcal{P}\}.$$

Finally, let $\tau(\mathcal{P})$ denote the smallest topology on $\cup\mathcal{P}$ containing $\mathcal{P}$. By definition, since $\tau(\mathcal{P})$ is a topology, it is closed under unions or finite intersections. The smallest topology containing $\mathcal{P}$ is the topology for which $\mathcal{P}$ is the subbase. More precisely, $\tau(\mathcal{P})$ is the topology generated by the base $I(\mathcal{P})$ (if $\mathcal{P}$ is finite — otherwise, $I(\mathcal{P})$ should be replaced by the collection of finite intersections). By definition, then, for every $p \in \cup\mathcal{P}$ and every $U \in \tau(\mathcal{P})$ such that $p \in U$, there exists $V \in I(\mathcal{P})$ such that $p \in V \subseteq U$.

(See Kelly [1955], pp. 46–48.) The natural alternatives to consider are $I(S)$, $U(S)$, and $\tau(S)$, which I will simply denote I, U, and τ respectively.

Remark 2. When Gilboa and Schmeidler [1992] show that Choquet expected utility is equivalent to standard expected utility on an enlarged state set, the enlarged state set they consider is I. (Their framework is different and so the construction is not described this way; however, it is not hard to show that it is equivalent.) Similarly, the equivalence of belief functions to additive functions on an enlarged state space, as shown by Shafer [1976] and others, uses I as the enlarged state space.

Rescher and Brandom call the states in $I \setminus S$ *schematic worlds* and those in $U \setminus S$ *superposed worlds.* They discuss at some length the interpretation of these two procedures and some of the properties of the deduction relation they imply. They do not discuss the worlds in τ and the following characterization of their logics is also new.

First, I present a lemma describing how the topology can be constructed from the U and I operators.

Lemma 1. *If $\mathcal{P}$ is finite, $\tau(\mathcal{P}) = U(I(\mathcal{P}))$.*

As the lemma should suggest, introducing the worlds in τ is not a major generalization of the Rescher and Brandom analysis. This point will be seen even more clearly in a moment.

I will now introduce some properties we might want the agent's logic to satisfy.

Definition. $\vdash_*$ preserves simple inference *(PSI) if $p \vdash q$ implies $p \vdash_* q$.*

Intuitively, if $\vdash_*$ preserves simple inference, then an agent learning the single premise p infers any one conclusion tautologically implied by p. Note in particular that $\top \vdash p$ for any tautology p. Since I require $\top \in s^*$ for all s^*, PSI implies that every tautology p is true in every world in S^*. In this sense, it requires that the agent knows all tautologies.

One might be puzzled by the reference to simple inferences. After all, the "one" premise p could actually be a conjunction of numerous propositions and the "one" conclusion q could be the disjunction of several propositions. Hence one might be tempted to conclude that an agent who satisfies PSI is, in fact, a perfect reasoner. As noted by Rescher and Brandom [1979] and hinted at in the discussion before the definition of $\vdash$, this view is not correct. To see the point most simply, suppose that $p \wedge p' \vdash q$. If PSI holds, an agent who learns $p \wedge p'$ will infer q. However, PSI does not rule out the possibility that an agent who already knows p and learns p' still fails to infer q. In other words, PSI does not imply that the agent combines the two premises p and p' to recognize $p \wedge p'$. Hence he may not infer q. This motivates the following additional property on reasoning.

Definition. $\vdash_*$ *satisfies* perfect conjunction *(PC) if: (i)* $p, q \vdash_* p \wedge q$, *(ii)* $p \wedge q \vdash_* p$, *and (iii)* $p \wedge q \vdash_* q$.

Note that if $\vdash_*$ satisfies PSI, then (*ii*) and (*iii*) are redundant.

Just as PSI does not guarantee that the agent combines multiple premises into their conjunction, it does not guarantee that he combines multiple conclusions into their disjunction. This suggests the following "dual" to PC:

Definition. $\vdash_*$ *satisfies* perfect disjunction *(PD) if (i)* $p \vdash_* p \vee q$, *(ii)* $q \vdash_* p \vee q$, *and (iii)* $p \vee q \vdash_* p, q$.

If $\vdash_*$ satisfies PSI, then (*i*) and (*ii*) are redundant.

The following result provides a characterization of logics which use the τ operator to construct new worlds. It shows that all logics which generalize Rescher and Brandom by using τ must satisfy PSI. Furthermore, *any* logic satisfying PSI is, in this sense, a generalization of Rescher and Brandom's logic of inconsistency.

Theorem 2. $\vdash_*$ *satisfies PSI iff* $S^* \subseteq \tau$.

The next result clarifies the implication of restricting the set of worlds further to either using only the schematic worlds or only the superposed worlds.

Theorem 3. $\vdash_*$ *satisfies PSI and PC iff* $S^* \subseteq I$. $\vdash_*$ *satisfies PSI and PD iff* $S^* \subseteq U$.

Corollary. $\vdash_*$ *satisfies PSI, PC, and PD iff* $\vdash_* = \vdash$ *and* $S = S^*$.

In other words, any logic satisfying PSI and PC must be constructed from Rescher and Brandom's schematic worlds, while any logic satisfying PSI and PD must be constructed from their superposed worlds. If we only impose PSI, the logic is a generalization of theirs, constructed from the worlds in τ. Given this, it is clear that introducing τ simply unifies these two logics.

At this point, an example may be useful. Suppose we have two atomic propositions, p and q. The other propositions all take the form $p \vee q$, $\neg p$, etc. Hence $S = \{s_1, s_2, s_3, s_4\}$ where

$$p, q \in s_1,$$

$$p, \neg q \in s_2,$$

$$\neg p, q \in s_3$$

and

$$\neg p, \neg q \in s_4.$$

Of course, the other propositions in s_1 are the tautologies and other implications of p and q, such as $\neg\neg p$ or $p \to q$, and analogously for the other states.

First, suppose that $S^* = \{s_1, s_2, s_3, s_4, s_1 \cup s_2\}$. Since $S^* \subset U$, we know from Theorem 3 that PSI and PD must hold. However, PC will *not* be satisfied. To see this explicitly, note that $q \in s_1$ and $\neg q \in s_2$. Hence both q and $\neg q$ are elements of $s_1 \cup s_2$. However, $q \wedge \neg q \notin s_i$ for any i, so $q \wedge \neg q \notin s_1 \cup s_2$. Hence $S^*(q \wedge \neg q) = \emptyset$, while $S^*(q) \cap S^*(\neg q) = s_1 \cup s_2$. By definition of $\vdash_*$, then, $q, \neg q \not\vdash_* q \wedge \neg q$, contradicting PC.

As an alternative, suppose $S^* = \{s_1, s_2, s_3, s_4, s_1 \cap s_2\}$. Now $S^* \subset I$, so that PSI and PC will hold. However, PD will not be satisfied. To see this, note that $\neg q \notin s_1$ and $q \notin s_2$, so that $\neg q, q \notin s_1 \cap s_2$. On the other hand, $q \vee \neg q$ is a tautology, so that it is contained in every s_i and therefore $q \vee \neg q \in s_1 \cap s_2$. Therefore, $S^*(q \vee \neg q) = S^*$, but $S^*(q) \cup S^*(\neg q) = S$. Since $S^* \not\subseteq S$, we see that $q \vee \neg q \not\vdash_* q, \neg q$.

As a last example, suppose $S^* = \{s_1, s_2, s_3, s_4, s_2 \cup s_4\}$. Just as in the first example, $S^* \subset U$, so that $\vdash_*$ must satisfy PSI and PD but will not satisfy PC. A particularly interesting example of this failure is the following. Recall that $p \to q$ is true whenever p and q are true or p is false. Hence $S(p \to q) = \{s_1, s_3, s_4\}$. Since $p \to q \in s_4$, we certainly have $p \to q \in s_2 \cup s_4$. Hence $S^*(p \to q) = \{s_1, s_3, s_4, s_2 \cup s_4\}$. Also, $S^*(p) = \{s_1, s_2, s_2 \cup s_4\}$. Hence

$$S^*(p \to q) \ \cap \ S^*(p) = \{s_1, s_2 \cup s_4\}.$$

Notice, though, that $q \notin s_2 \cup s_4$. Hence $p, p \to q \not\vdash_* q$ — that is, if the agent learns that $p \to q$ and also learns that p, he does not infer q. On the other hand,

$$S^*(p \wedge (p \to q)) = \{s_1\}$$

as $p \wedge (p \rightarrow q) \notin s_2 \cup s_4$. Hence an agent who *simultaneously* learns p and $p \rightarrow q$ does infer q. This is simply an implication of PSI since $p \wedge (p \rightarrow q) \vdash q$.

I now characterize preferences such that we can attribute a form of the logic of inconsistency to the agent.

In light of Theorem 2, if we can represent $\{\succ_p\}$ by extended expected utility where $S^* \subseteq \tau$, then the implied logic satisfies PSI. That is, S^* will have the property that if $S(p) \subseteq S(q)$, then $S^*(p) \subseteq S^*(q)$. Clearly, then, if $S(p) = S(q)$, we must have $S^*(p) = S^*(q)$. Hence if $\{\succ_p\}$ is XEUR with PSI, it must satisfy RLO. In fact, RLO plus representability is sufficient for $\{\succ_p\}$ to be XEUR with PSI. In this sense, revealed logical omniscience is exactly the preference implication of PSI.

Theorem 4. *$\{\succ_p\}$ is XEUR with $\vdash_*$ satisfying PSI iff $\{\succ_p\}$ satisfies representability and RLO.*

One interesting aspect of Theorem 4 is that RLO — a property which only refers to logical *equivalence* — tells us that we can treat the agent as correctly recognizing all simple logical *implications*. Even though RLO would seem to allow the possibility that the agent correctly recognizes equivalence but occasionally makes errors regarding implications, Theorem 4 implies that this distinction has no behaviorial content. On the other hand, it is possible, at least in principle, that the distinction could be important if one wishes to find a representation that satisfies certain other properties as well.

Suppose now that $\{\succ_p\}$ is extended expected utility representable with PSI and PC — that is, with $S^* \subseteq I$. Note that $S^* = I$ implies

$$S^*(p) = \{s \in 2^P \mid s = \cap B, \text{ some } B \subseteq S(p)\}.$$

Hence if $S(p)$ is a singleton, $S^*(p) = S(p)$. To see the implication of this, suppose $S(p) = \{s\}$. Suppose $\{\succ_p\}$ is XEUR with $S^* = I$, utility function

u, and measure μ. Then if $\mu(s) > 0$,

$$\mathrm{E}_{\mu}[u(h(f)(s^*)) \mid s^* \in S^*(p)] = u(f(s)).$$

Therefore, it must be true that $f \succ_p g$ iff $u(f(s)) > u(g(s))$. Furthermore, the same must be true for *any* p such that $S(p)$ is a singleton. Put differently, u must represent preferences over consequences for *every* state. This kind of state independence is also a necessary condition for EUR and is an obvious weakening of Savage's [1954] P3.

To state this condition more formally:

Definition. $\{\succ_p\}$ *satisfies* weak state independence *(WSI) if there exists* $u : X \to \mathbf{R}$ *such that for all* $s \in S$ *and every nonnull* p *such that* $S(p) = \{s\}$, $f \succ_p g$ *iff* $u(f(s)) > u(g(s))$.

The argument above shows that this condition is necessary for $\{\succ_p\}$ to be XEUR with $S^* \subseteq I$. I show below that representability, RLO, and WSI are sufficient with one auxiliary assumption. In this sense, the preference implication of PC is WSI. The auxiliary assumption is

(A2) $\{\succ_p\}$ satisfies WSI where u is onto.

This assumption implies that there are no best or worst consequences (where "best" and "worst" are defined relative to u) and u has no "gaps."

Theorem 5. *If* $\{\succ_p\}$ *is XEUR with* $\vdash_*$ *satisfying PSI and PC, then* $\{\succ_p\}$ *must satisfy representability, RLO, and WSI. If representability, RLO, WSI, and (A2) hold, then* $\{\succ_p\}$ *is XEUR with* $\vdash_*$ *satisfying PSI and PC.*

Finally, I turn to the preference conditions under which there is a representation where $\vdash_*$ satisfies PSI and PD. The additional preference requirement for PD is analogous to but more complex than the condition for PC. To see it intuitively, suppose we have such a representation with μ, u, and h denoting the relevant functions. Fix any state s and a proposition p such that $S(p) = S \setminus \{s\}$. We know from Theorem 3 that satisfying PD requires $S^* \subseteq U$. Suppose $S^* = U$. Then

$$\mathrm{E}_\mu[u(h(f)(s^*))] = \sum_{s \in S} \mu(s) u(f(s)) + \sum_{B \subseteq S, \# B \geq 2} \mu(\cup B) u(h(f)(\cup B)).$$

Let $u_\top$ denote the function on the right–hand side. Also,

$$\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] = \sum_{s' \neq s} \frac{\mu(s')}{1-\mu(s)} u(f(s'))$$
$$+ \sum_{B \subseteq S, \# B \geq 2} \frac{\mu(\cup B)}{1-\mu(s)} u(h(f)(\cup B)).$$

To see this, note that $p \in s'$ for all $s' \neq s$. Hence for any $B \subseteq S$ such that $B \neq \{s\}$, $p \in \cup B$. Let u_p denote the function on the right–hand side. Note that

$$u_\top(f) = [1-\mu(s)] u_p(f) + \mu(s) u(f(s)).$$

This must hold for every such s, p, and f. Notice that the function $[1 - \mu(s)]u_p(f)$ represents the same preferences as u_p. Hence this condition implies that we must be able to find a representation of $\succ$, say $u_\top$, and, for every s and every p with $S(p) = S \setminus \{s\}$, a representation of $\succ_p$, say $\hat{u}_p$, such that

$$u_\top(f) - \hat{u}_p(f) = \mu(s) u(f(s)).$$

In a sense, this condition is the "dual" to WSI. Weak state independence required us to find a function $u(x)$ that represented preferences conditional on a single state. Equivalently, we needed to find functions $\mu(s)$ and $u(x)$ such that the product represented preferences given a single state. This condition requires us to find functions $\mu(s)$ and $u(x)$ such that the product represents the *difference* in utility associated with a single state.

To state this condition formally,

Definition. $\{\succ_p\}$ *satisfies* dual weak state independence *(DWSI) if there exists* $u_\top$ *representing* $\succ$ *and, for each* $s \in S$ *and each* p *with* $S(p) = S \setminus \{s\}$, *a function* u_p *representing* $\succ_p$ *such that there exists a function* $\mu : S \to \mathbf{R}_+$ *and* $u : X \to \mathbf{R}$ *such that*

$$f(s) = x, \; S(p) = S \setminus \{s\} \implies \mu(s)u(x) \;=\; u_\top(f) - u_p(f).$$

As the discussion above shows, DWSI is necessary for the existence of an XEUR representation where $\vdash_*$ satisfies PSI and PD. To demonstrate sufficiency, as in the case of Theorem 5, it is convenient to strengthen the condition to:

(A3 $\{\succ_p\}$ satisfies DWSI where u is onto.

Theorem 6. *If* $\{\succ_p\}$ *is XEUR with* $\vdash_*$ *satisfying PSI and PD, then* $\{\succ_p\}$ *must satisfy representability, RLO, and DWSI. If representability, RLO, DWSI, and (A3) hold, then* $\{\succ_p\}$ *is XEUR with* $\vdash_*$ *satisfying PSI and PD.*

Summarizing, then, representability is the preference implication of XEUR, RLO is the additional preference implication of PSI, and WSI (DWSI) is the additional preference implication of PC (PD).

V. Conclusion.

It bears emphasizing that the analysis here is still very preliminary. There are numerous open questions that I am engaged in trying to answer. A few of these are sketched here.

First, one obvious direction for further work involves continuing along the lines above. There are certainly many potentially interesting logics that could be studied in a similar manner to my analysis of Rescher and Brandom's logic of inconsistency. In particular, it would be quite interesting to see how a modal logic could be derived.

An issue not addressed here which I discuss at some length in Lipman [1992a] is the agent's understanding of the available acts. The extension function h describes the agent's perception of the way the outcomes of possible actions depend on what is true about the world. In Lipman [1992a], I relate conditions on preferences to various statements about how accurately the agent understands the available acts. There are still many open questions in this direction as well.

In addition, there are a number of seemingly related ideas in the literature which may relate to and/or cast light on the results here, some of which were mentioned in Remark 2. Also, there may be other interpretations of the model here related to suggestions in the literature. For example, Stalnaker [1987] suggests that imperfectly reasoning agents might be thought of as collections of logically closed belief states. Is there a formal way to interpret XEUR as modeling this kind of phenomena?

Finally, the purpose of this project is to generate a useful model of bounded rationality for economics and game theory. The analysis carried out so far and the further possibilities described above are focused on understanding the structure of the model. Ultimately, the usefulness of the model will only be demonstrated by trying to apply it. This was one of the purposes of Lipman [1992b], but much further work in this direction is still needed.

APPENDIX

Proof of Theorem 1.

The proof is by construction. Define an equivalence relation $\cong$ on $\bar{P}$ by $p \cong q$ if $\succ_p = \succ_q$ and $S(p) = S(q)$. (It is easy to see that this is an equivalence relation.) Let $\mathcal{D} = \{D_0, \ldots, D_m\}$ denote the partition of $\bar{P}$ induced by $\cong$. Note that (A1) implies that $\mathcal{D}$ is finite. For any proposition $p \in \bar{P}$, let $D(p)$ denote the event in $\mathcal{D}$ which contains p. For notational convenience, let $D_0 = D(\top)$. For the moment, assume that $D_0 \cup D_k \notin S$ for any k.

Let

$$S^* = S \cup \{s \in 2^P \mid s = D_0 \cup D_k \text{ for some } k\}.$$

The key fact to observe about this construction is that for any $p \in \bar{P}$ such that $D(p) \neq D_0$, there is exactly one $s^* \in S^* \setminus S$ such that $p \in s^*$. Hence for such a p

$$S^*(p) = S(p) \cup \{D_0 \cup D(p)\}.$$

If $D(p) = D_0$, then $p \in s^*$ for every $s^* \in S^*$. (By construction, if $D(p) = D_0$, then p is a tautology and so is in every $s \in S$.) Hence for such a p, $S^*(p) = S^*$.

Let $u(x) = x$ and $\mu(s^*) = 1/\#S^*$ for all $s^* \in S^*$ (where for any set B, $\#B$ denotes the cardinality of B). Fix a collection of functions u_p, $p \in \bar{P}$ satisfying representability. Without loss of generality, assume that $p \cong q \implies u_q = u_p$. Also, for each $D_k \in \mathcal{D}$, fix a proposition $p_k \in D_k$. Extend the lotteries to S^* as follows. Of course, if $s^* \in S$, then $h(f)(s^*) = f(s^*)$. If $s^* = D_0 \cup D_k$ for $k \neq 0$, then

$$h(f)(s^*) = u_{p_k}(f) - \sum_{s \in S(p_k)} f(s).$$

Finally, for $s^* = D_0$,

$$h(f)(s^*) = u_{\top}(f) - \sum_{k=1}^{m} [u_{p_k}(f) - \sum_{s \in S(p_k)} f(s)] - \sum_{s \in S} f(s).$$

Since $X = \mathbf{R}$ and the utility functions are all real–valued, this extension is legitimate.

For any $p \in \bar{P}$ such that $D(p) \neq D_0$ and any $f \in F$,

$$\begin{aligned}
\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] &= \frac{1}{\#S^*(p)} \sum_{s^* \in S^*(p)} h(f)(s^*) \\
&= \frac{1}{\#S^*(p)} \left[u_p(f) - \sum_{s \in S(p)} f(s) + \sum_{s \in S(p)} f(s) \right] \\
&= \frac{1}{\#S^*(p)} u_p(f).
\end{aligned}$$

Clearly, then, this function will represent $\succ_p$ since u_p represents $\succ_p$. For any $p \in \bar{P}$ with $D(p) = D_0$ and any $f \in F$,

$$\begin{aligned}
\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] &= \mathrm{E}_\mu[h(f)(s^*)] = \frac{1}{\#S^*} \sum_{s^* \in S^*} h(f)(s^*) \\
&= \frac{1}{\#S^*} \Bigg[u_\top(f) - \sum_{k=1}^{m} [u_{p_k}(f) - \sum_{s \in S^*(p_k)} f(s)] - \sum_{s \in S} f(s) \\
&\quad + \sum_{k=1}^{m} [u_{p_k}(f) - \sum_{s \in S(p_k)} f(s)] + \sum_{s \in S} f(s) \Bigg] \\
&= \frac{1}{\#S^*} u_\top(f).
\end{aligned}$$

Again, this function obviously represents $\succ_p$ since it represents $\succ_\top$ and $\succ_\top = \succ_p$ for all p with $\bar{S}_p = \bar{S}_\top$.

All that remains is to deal with the case where $D_0 \cup D(p) \in S$ for some p. First, suppose that $D_0 \in S$. This means that there is some $s \in S$ such that

$$s = \{p \in \bar{P} \mid S(p) = S \text{ and } \succ_p = \succ\}.$$

Note that this requires every proposition in s to be a tautology. Suppose there is any $q \in \bar{P}$ such that $q \notin s$. Since s is a logically consistent world, $\neg q \in s$. But since every proposition in s is a tautology, $\neg q \in s'$ for all $s' \in S$, so $q \notin s'$ for any $s' \in S$, contradicting the assumption that $q \in \bar{P}$.

Hence $s = \bar{P}$. But then $\succ_p = \succ$ for every $p \in \bar{P}$. Representability trivially implies $\{\succ_p\}$ is XEUR in this case.

So suppose that $D_0 \cup D(p) \in S$ for some $p \notin D_0$. Suppose $D(p) = \{p\}$. Since every $q \in D_0$ is a tautology, this means there is a state $s \in S$ which has exactly one proposition, p, which is not a tautology. Consider the proposition $\neg\neg p$. By the definition of a logically consistent state, $\neg\neg p \in s$. But if p is not a tautology, $\neg\neg p$ cannot be a tautology either, contradicting the hypothesis that p is the only proposition which is not a tautology in s.

Hence for every $p \notin D_0$ such that $D_0 \cup D(p) \in S$, we must have $\#D(p) \geq 2$. Hence we can simply partition all such $D(p)$ into nonempty sets $D^1(p)$ and $D^2(p)$. It is easy to see that we cannot have $D^i(p) \cup D_0 \in S$ for $i = 1, 2$. It is straightforward to repeat the construction above where S^* consists of S plus the various sets $D^i(p) \cup D_0$. ▌

Proof of Lemma 1.

By definition, a topology is closed under (finite) intersections and unions. Hence $U(I(\mathcal{P})) \subseteq \tau(\mathcal{P})$. But consider any set $U \in \tau(\mathcal{P})$. By definition, for each $p \in U$, there must be $V_p \in I(\mathcal{P})$ such that $p \in V_p \subseteq U$. But then

$$\bigcup_{p \in U} V_p = U,$$

so $U \in \tau(\mathcal{P})$ implies $U \in U(I(\mathcal{P}))$. Hence $\tau(\mathcal{P}) \subseteq U(I(\mathcal{P}))$ so that $\tau(\mathcal{P}) = U(I(\mathcal{P}))$.[4] ▌

Proof of Theorem 2.

First, suppose that $S^* \subseteq \tau$. I show if $p \vdash q$ (*i.e.*, $S(p) \subseteq S(q)$), then $p \vdash_* q$ ($S^*(p) \subseteq S^*(q)$). Consider any $s^* \in \tau$ such that $p \in s^*$. Since I is

[4] If $\mathcal{P}$ is not finite, all statements here are still true where we simply replace $I(\mathcal{P})$ with the set of all finite intersections of sets in $\mathcal{P}$.

a base for τ, there exists V such that $p \in V \subseteq s^*$ where $V = \cap B$, some $B \subseteq S$. Hence $p \in s$ for all $s \in B$, so $B \subseteq S(p)$. But then $B \subseteq S(q)$, so we must have $q \in s$ for all $s \in B$. Hence $q \in V$, so $q \in s^*$. Hence $S^*(p) \subseteq S^*(q)$.

To show the converse, suppose $\vdash_*$ satisfies PSI so that $S(p) \subseteq S(q)$ implies $S^*(p) \subseteq S^*(q)$. In particular, this implies that if $S(p) = S(q)$, we must have $S^*(p) = S^*(q)$. For each nonempty $B \subseteq S$, let

$$P_B = \{p \in \bar{P} \mid S(p) = B\}.$$

Obviously, every $p \in \bar{P}$ is in exactly one P_B. Fix any $s^* \in S^*$. Suppose that there is a $p \in P_B$ such that $p \in s^*$. For any $q \in P_B$, $S(p) = S(q)(= B)$, so PSI implies $S^*(p) = S^*(q)$. Hence $s^* \in S^*(p)$ implies $q \in s^*$ for all $q \in P_B$. Therefore, PSI implies that every s^* can be written in the form

$$s^* = \bigcup_{B \in \mathcal{B}(s^*)} P_B$$

where $\mathcal{B}(s^*) = \{B \subseteq S \mid P_B \subseteq s^*\}$. Also, suppose $B \in \mathcal{B}(s^*)$ and $B \subseteq B'$. Let $p \in P_B$ and $q \in P_{B'}$, so $S(p) \subseteq S(q)$. Then since $s^* \in S^*(p)$, PAI requires $s^* \in S^*(q)$ — that is, $q \in s^*$ or $B' \in \mathcal{B}(s^*)$. In short,

$$(2) \qquad B \in \mathcal{B}(s^*),\ B \subseteq B' \implies B' \in \mathcal{B}(s^*)$$

Note that PSI also implies that if $S(p) = \emptyset = S(\perp)$, then $S^*(p) = S^*(\perp)$. By assumption, $S^*(\perp) = \emptyset$. Hence for all $s^* \in S^*$, $\emptyset \notin \mathcal{B}(s^*)$.

For any $\mathcal{B} \subseteq 2^S$, let

$$\min \mathcal{B} = \{B \in \mathcal{B} \mid \nexists B' \in \mathcal{B} \text{ with } B' \subset B\}.$$

Since S is finite, every $\mathcal{B}$ is finite and so $\min \mathcal{B} \neq \emptyset$. Clearly, for any $s^* \in S^*$, (2) implies

$$s^* = \bigcup_{B \in \min \mathcal{B}(s^*)} \bigcup_{B' \supseteq B} P_{B'}.$$

But for any $B \subseteq S$,

$$\bigcup_{B' \supseteq B} P_{B'} = \{p \in \bar{P} \mid B \subseteq S(p)\}.$$

Hence

$$p \in \bigcup_{B' \supseteq B} P_{B'} \iff p \in s,\ \forall s \in B \iff p \in \cap B.$$

That is,

$$\bigcup_{B' \supseteq B} P_{B'} = \cap B.$$

Hence every s^* is a union of elements of I. By Lemma 1, $\tau = U(I)$, so every $s^* \in \tau$. ▌

Proof of Theorem 3.

The proof makes use of the following lemma.

Lemma 2. *For every $B \subseteq S$, there exists $p_B \in P$ such that $S(p_B) = B$.*

Proof: If $s \neq s'$, then there must be $p \in s$ such that $p \notin s'$. For each s and s' with $s \neq s'$, let $p(s, s')$ denote such a proposition. Then

$$S\left(\bigwedge_{s' \neq s} p(s, s')\right) = \{s\},$$

where $\bigwedge_{s' \neq s}$ denotes the conjunction of the propositions. Hence for every singleton $\{s\}$, there is a proposition p_s such that $S(p) = \{s\}$. For nonsingleton sets B, let

$$p_B = \bigvee_{s \in B} p_s.$$

It is easy to see that $S(p_B) = B$. Finally, for $B = \emptyset$, let $p_B = p \wedge \neg p$ for any proposition p. ▌

For the proof of the theorem, first suppose $S^* \subseteq I$ ($\subseteq \tau$). By Theorem 2, PSI must hold. Hence conditions (*ii*) and (*iii*) of the definition of PC must hold, so that we only need to show that $p, q \vdash_* p \wedge q$. So consider any $s^* \in S^*(p) \cap S^*(q)$. By the definition of I, there exists $B \subseteq S$ such that $s^* = \cap B$ and $p, q \in s$ for all $s \in B$. Hence $p \wedge q \in s$ for all $s \in B$, so $p \wedge q \in s^*$. Hence $S^* \subseteq I$ implies $p, q \vdash_* p \wedge q$, so PC holds.

Similarly, if $S^* \subseteq U$, PSI must hold. Since $p \vdash p \vee q$ and $q \vdash p \vee q$, conditions (i) and (ii) of the definition of PC are automatically satisfied. So we only need to show that $p \vee q \vdash_* p, q$. For every $s^* \in U$, if $s^* \in S^*(p \vee q)$, there is $s \in S$ such that $p \vee q \in s \subseteq s^*$. Hence either $p \in s$ or $q \in s$ so either $p \in s^*$ or $q \in s^*$ so $s^* \in S^*(p) \cup S^*(q)$. Therefore, $S^* \subseteq U$ implies $p \vee q \vdash_* p, q$ or PD.

For the converses, first suppose S^* satisfies PSI and PC. By Theorem 2, we know that each s^* can be written in the form

$$s^* = \bigcup_{B \in \mathcal{B}} P_B$$

where $P_B = \{p \in P \mid S(p) = B\}$ and $\mathcal{B}$ is a collection of subsets of S. Suppose $p \in P_B$ and $q \in P_{B'}$ where $B,\ B' \in \mathcal{B}$. Then $s^* \in S^*(p) \cap S^*(q)$, so PC implies $s^* \in S^*(p \wedge q)$. Hence $p \wedge q \in s^*$, or $S(p \wedge q) \in \mathcal{B}$. But $S(p \wedge q) = S(q) \cap S(q) = B \cap B'$. Hence if $B,\ B' \in \mathcal{B}$, we must have $B \cap B' \in \mathcal{B}$. By the finiteness of S, then, $\hat{B} \equiv \cap \mathcal{B} \in \mathcal{B}$. This implies that

$$\min \mathcal{B} = \{\hat{B}\},$$

so that the same argument as in the proof of Theorem 2 shows that $s^* = \cap \hat{B}$, or $s^* \in I$.

Suppose, then, that PSI and PD hold. Again using Theorem 2, we know that every s^* can be written in the form

$$s^* = \bigcup_{B \in \min \mathcal{B}} \cap B.$$

Suppose there exists $B \in \min \mathcal{B}$ such that $\#B \geq 2$. Let B_1 and B_2 be a partition of B with $B_i \neq \emptyset$, $i = 1, 2$. Let $S(p) = B_1$ and $S(q) = B_2$. Then

$S(p \vee q) = B$. Hence $p \vee q \in s^*$. However, $p \notin s^*$ and $q \notin s^*$, contradicting PD. Hence every $B \in \min \mathcal{B}$ is a singleton, so $s^* \in U$. ▌

Proof of Theorems 4 and 5.

The following lemma is useful.

Lemma 3. *For every $B_1, B_2 \subseteq S$ with $B_1 \neq B_2$, $\cap B_1 \neq \cap B_2$.*

Proof: Using the notation of Lemma 2, clearly, $p_{B_i} \in \cap B_i$, $i = 1, 2$. However, $p_{B_1} \in \cap B_2$ iff $B_2 \subseteq B_1$. Since $B_1 \neq B_2$, it cannot be true that both B_1 and B_2 are subsets of the other. Hence either $p_{B_1} \notin \cap B_2$ or $p_{B_2} \notin \cap B_1$. ▌

For convenience, I first prove Theorem 5. The discussion in the text proves the first sentence of the theorem. To prove the second, note that by Theorem 3, it is sufficient to show that $\{\succ_p\}$ is XEUR with $S^* = I$. Fix u satisfying (A2). For simplicity, assume every proposition p such that $\#S(p) = 1$ is nonnull. (It is straightforward to adapt the proof to cover any null p with $S(p)$ a singleton.) Let $\mu(s^*) = 1/\#S^*$ for all $s^* \in S^* = I$. For each p such that $\#S(p) \geq 2$, fix u_p satisfying representability. Without loss of generality, assume that $u_p = u_q$ if $S(p) = S(q)$. (Recall that by RLO, $S(p) = S(q)$ implies $\succ_p = \succ_q$.) For each $B \subseteq S$, fix a proposition, p_B, such that $S(p_B) = B$. By Lemma 2, this is possible. Fix any function $v : \mathbf{R} \to \mathbf{R}$ such that $u(v(x)) = x$. That is, v is u^{-1} or a selection from u^{-1} if the inverse is not single–valued. By (A2), $v(x)$ exists for all $x \in \mathbf{R}$.

Construct h as follows. Of course, $h(f)(s) = f(s)$ for all $s \in S$. For any s^* such that $s^* = \cap B$ with $B \subseteq S$, $\#B = 2$, let

$$h(f)(s^*) = v\left(u_{p_B}(f) - \sum_{s \in B} u(f(s))\right).$$

(By Lemma 3, given any $s^* \in S^*$, there is exactly one $B \subseteq S$ such that $s^* = \cap B$.) In general, for s^* such that $s^* = \cap B$, $B \subseteq S$ with $\#B = k \geq 3$, let

$$h(f)(s^*) = v\left(u_{p_B}(f) - \sum_{\ell=2}^{k-1} \sum_{B' \subset B, \# B' = \ell} u(h(f)(\cap B')) - \sum_{s \in S(p_B)} u(f(s))\right).$$

Consider any $p \in \bar{P}$ and any $f \in F$. If $S(p) = \{s\}$, then

$$\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] = u(f(s)).$$

By WSI, this represents $\succ_p$. Suppose, then, that $\#S(p) \geq 2$. Then

$$\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] = \frac{1}{\#S^*(p)} \sum_{s \in S^*(p)} u(h(f)(s^*))$$

$$= \frac{1}{\#S^*(p)} \sum_{B \subseteq S(p)} u(h(f)(\cap B))$$

$$= \frac{1}{\#S^*(p)} \left[u[h(f)(\cap S(p))] + \sum_{\ell=1}^{\#S(p)-1} \sum_{B' \subset S(p), \# B' = \ell} u(h(f)(\cap B')) + \sum_{s \in S(p)} u(f(s)) \right]$$

$$= \frac{1}{\#S^*(p)} u_p(f)$$

which represents $\succ_p$ by assumption. This completes the proof of Theorem 5.

Intuitively, WSI is necessary for Theorem 5 because if $S(p)$ is a singleton, $S^* = I$ makes $S^*(p) = S(p)$. Hence in conditioning on such a p, we cannot use the extension to help represent the preferences. As this intuition suggests, if we add just a few more states to S^*, WSI is no longer necessary. In fact, if we continue to assume RLO, we can choose these additional states from τ. This is what gives us the proof of Theorem 4, as follows.

Again, the first sentence of the theorem is proven in the text. To prove the second sentence, first note that, by Theorem 2, it suffices to show that

$\{\succ_p\}$ is XEUR with $S^* \subseteq \tau$. Fix any $s \in S$ and refer to this state as s_1. Let

$$S^*_{U1} = \{\cup S\} \bigcup \{s^* \in 2^P \mid s^* = s_1 \cup s, \text{ some } s \in S\}$$

and let $S^* = I \cup S^*_{U1}$. Clearly, then $S^* \subseteq \tau$.

Let $u(x) = x$ and let $\mu(s^*) = 1/\#S^*$. Let $n = \#S$. Fix u_p, $p \in \bar{P}$ satisfying representability where $u_p = u_q$ whenever $\succ_p = \succ_q$. As before, for each $B \subseteq S$, fix p_B such that $S(p_B) = B$. Note that

$$S^*(p_B) = \{s^* \in S^* \mid s^* \in B \text{ or } s^* = \cap B$$

$$\text{or } s^* \in S^*_{U1} \text{ and } s \subseteq s^* \text{ some } s \in B\}.$$

Define h as follows. Of course, for any $s \in S$, $h(f)(s) = f(s)$. For any $s \neq s_1$, let

$$h(f)(s \cup s_1) = u_{p_s}(f) - f(s) - h(f)(\cup S).$$

Let

$$h(f)(\cup S) = \frac{1}{n-2}\left[f(s_1) - u_{p_{s_1}}(f) - \sum_{s \neq s_1}\left(u_{p_s}(f) - f(s)\right)\right].$$

For $B \subseteq S$ with $\#B = 2$, let

$$h(f)(\cap B) = u_{p_B}(f) - \sum_{s \in B} f(s) - \sum_{\{s^* \in S^*_{U1} \mid \exists s \subseteq s^* \text{ with } s \in B\}} h(f)(s^*).$$

Finally, for $B \subseteq S$ with $\#B = k \geq 3$,

$$h(f)(\cap B) = u_{p_B}(f) - \sum_{s \in B} f(s) - \sum_{\ell=2}^{k-1} \sum_{B' \subset B, \#B' = \ell} h(f)(\cap B') - \sum_{\{s^* \in S^*_{U1} \mid \exists s \subseteq s^* \text{ with } s \in B\}} h(f)(s^*).$$

Consider any p such that $S(p)$ is a singleton, say $\{s\}$. If $s \neq s_1$,

$$\begin{aligned} \mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] &= \frac{1}{3}\left[f(s) + h(f)(s \cup s_1) + h(f)(\cup S)\right] \\ &= \frac{1}{3} u_p(f), \end{aligned}$$

which obviously represents $\succ_p$. If $s = s_1$,

$$\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] = \frac{1}{n}\left[f(s_1) + \sum_{s \neq s_1} h(f)(s \cup s_1) + h(f)(\cup S)\right]$$

$$= \frac{1}{n}\left[f(s_1) + \sum_{s \neq s_1} \left(u_{p_s}(f) - f(s) - h(f)(\cup S)\right) + h(f)(\cup S)\right]$$

$$= \frac{1}{n}\left[f(s_1) + \sum_{s \neq s_1} \left(u_{p_s}(f) - f(s)\right) - (n-2)h(f)(\cup S)\right]$$

$$= \frac{1}{n} u_p(f),$$

which again represents $\succ_p$. The argument for p with $\#S(p) \geq 2$ parallels the proof of Theorem 5 above. ▌

Proof of Theorem 6.

The following lemma is useful.

Lemma 4. *For every $B_1, B_2 \subseteq S$ with $B_1 \neq B_2$, $\cup B_1 \neq \cup B_2$.*

Proof: Using the notation of Lemma 2, clearly, $p_{S \setminus B_i} \notin \cup B_i$, $i = 1, 2$. However, $p_{S\setminus B_1} \in \cup B_2$ iff $B_2 \cap S \setminus B_1 \neq \emptyset$ — or $B_2 \not\subseteq B_1$. Since $B_1 \neq B_2$, it cannot be true that both B_1 and B_2 are subsets of the other. Hence either $p_{S\setminus B_1} \in \cup B_2$ or $p_{S \setminus B_2} \in \cup B_1$. ▌

The first sentence of the theorem is proved in the text. To prove the second, Theorem 3 implies that it is sufficient to show that the conditions imply an XEUR representation with $S^* = U$. Fix u and μ satisfying DWSI and (A3). Without loss of generality, assume that $\sum_{s \in S} \mu(s) < 1$. (Note that we can always replace $u_\top$ and each u_p with the same functions divided by a fixed constant and the new μ will be the old μ divided by this constant.) Let $\sum_{s \in S} \mu(s) = \beta$. Let

$$\alpha = \frac{1 - \beta}{\#(S^* \setminus S)}$$

and let $\mu(s^*) = \alpha$ for all $s^* \in S^* \backslash S$. It is easy to see that $\sum_{s^* \in S^*} \mu(s^*) = 1$.

For each p, fix u_p satisfying representability. Without loss of generality, assume that $u_p = u_q$ if $S(p) = S(q)$. For each $B \subseteq S$, fix a proposition, p_B, such that $S(p_B) = B$. By Lemma 2, this is possible. Fix any function $v : \mathbf{R} \to \mathbf{R}$ such that $u(v(x)) = x$. That is, v is u^{-1} or a selection from u^{-1} if the inverse is not single–valued. By (A3), $v(x)$ exists for all $x \in \mathbf{R}$. For each $p \in \bar{P}$ and each f, define

$$z_p(f) = u_p(f) - \sum_{s \in S(p)} \mu(s) u(f(s)).$$

Also, for p such that $S(p) = \emptyset$, let $z_p(f) = 0$ for all f.

Define $h(f)$ as follows. Of course, for $s \in S$, $h(f)(s) = f(s)$. For any $B \subseteq S$ such that $\#B = 2$, let

$$h(f)(\cup B) = v\left(\frac{1}{\alpha}[z_{\top}(f) - z_{\neg p_B}(f)]\right).$$

(By Lemma 4, given any $s^* \in S^*$, there is exactly one $B \subseteq S$ such that $s^* = \cup B$.) For any $B \subseteq S$ such that $\#B = k \geq 3$, let

$$h(f)(\cup B) = v\left(\frac{1}{\alpha}[z_{\top}(f) - z_{\neg p_B}(f)] - \sum_{B' \subset B, \#B' \geq 2} u(h(f)(\cup B'))\right).$$

The recursivity of the definition guarantees that $h(f)$ is well–defined.

Consider any $p \in \bar{P}$. Then

$$\begin{aligned} \mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)] = & \frac{1}{\mu(S^*(p))}\Bigg[\sum_{s \in S(p)} \mu(s) u(f(s)) \\ & + \alpha \sum_{B \subseteq S, \#B \geq 2, B \cap S(p) \neq \emptyset} u(h(f)(\cup B))\Bigg]. \end{aligned} \tag{3}$$

By definition,

$$\sum_{s \in S(p)} \mu(s) u(f(s)) = u_p(f) - z_p(f).$$

Also, note that

$$\sum_{B\subseteq S,\#B\geq 2,B\cap S(p)\neq\emptyset} u(h(f)(\cup B) = \sum_{B\subseteq S,\#B\geq 2} u(h(f)(\cup B)) - \sum_{B\subseteq S,\#B\geq 2,B\cap S(p)=\emptyset} u(h(f)(\cup B)).$$

Since $S \setminus S(p) = S(\neg p)$, we have $B \cap S(p) = \emptyset$ iff $B \subseteq S(\neg p)$. Hence the right–hand side of (3) can be rewritten as $1/\mu(S^*(p))$ times

$$\begin{aligned} u_p(f)-z_p(f) \ + \ \alpha u(h(f)(\cup S)) \ + \ \alpha \sum_{B\subset S,\#B\geq 2} u(h(f)(\cup B)) \\ - \ \alpha u(h(f)(\cup S(\neg p)) \ - \ \alpha \sum_{B\subset S(\neg p),\#B\geq 2} u(h(f)(\cup B)) \\ = u_p(f) - z_p(f) + z_\top(f) - 0 - [z_\top(f) - z_p(f)] \\ = u_p(f). \end{aligned}$$

Therefore, $\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(p)]$ represents $\succ_p$ as required. ▌

REFERENCES

Binmore, K., "Modeling Rational Players: Part I," *Economics and Philosophy*, 3, 1987, pp. 179–214.

Chew, S., and E. Karni, "Choquet Expected Utility with Finite State Space: Commutativity and Act Independence," University of California at Irvine working paper, 1992.

Fagin, R., and J. Halpern, "Belief, Awareness, and Limited Reasoning: Preliminary Report," IBM Research Laboratory working paper, April 1985.

Fagin, R., J. Halpern, and M. Vardi, "A Nonstandard Approach to the Logical Omniscience Problem," in R. Parikh, ed., *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Third Conference*, San Mateo: Morgan Kaufmann Publishers, 1990.

Gilboa, I., and D. Schmeidler, "Additive Representations of Non–Additive Measures and the Choquet Integral," Northwestern University working paper, 1992.

Gul, F., "Savage's Theorem with a Finite Number of States," Graduate School of Business, Stanford University working paper, 1992.

Hintikka, J., "Impossible Possible Worlds Vindicated," *Journal of Philosophical Logic*, 4, 1975, pp. 475–484.

Kelly, J., *General Topology*, New York: Springer–Verlag, 1955.

Lipman, B., "Decision Theory with Impossible Possible Worlds," Queen's University working paper, 1992a.

Lipman, B. L., "Limited Rationality and Endogenously Incomplete Contracts," Queen's University working paper, 1992b.

Lipman, B., "Information Processing and Bounded Rationality: A Survey," Queen's University working paper, 1993.

Morris, S., "Revising Knowledge: A Decision Theoretic Approach," University of Pennsylvania working paper, 1992.

Rantala, V., "Urn Models: A New Kind of Non–Standard Model for First–Order Logic," *Journal of Philosophical Logic*, 4, 1975, pp. 455–474.

Reny, P., "Rationality, Common Knowledge, and the Theory of Games," Ph.D. dissertation, Princeton University, 1986.

Rescher, N., and R. Brandom, *The Logic of Inconsistency*, Oxford: Basil Blackwell, 1979.

Rosenthal, R., "Games of Perfect Information, Predatory Pricing, and the Chain–Store Paradox," *Journal of Economic Theory*, 25, 1981, pp. 92–100.

Savage, L. J., *The Foundations of Statistics*, New York: Dover, 1954.

Shafer, G., *A Mathematical Theory of Evidence*, Princeton: Princeton University Press, 1976.

Shoesmith, D. J., and T. J. Smiley, *Multiple Conclusion Logic*, Cambridge: Cambridge University Press, 1978.

Stalnaker, R., *Inquiry*, Cambridge, MA: MIT Press, 1987.

van Damme, E., "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48, August 1989, pp. 476–496.