Decision Theory without Logical Omniscience: Toward an Axiomatic Framework for Bounded Rationality[1]

Barton L. Lipman

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario N6A 5C2

email: blipman@julian.uwo.ca

Current Draft

April 1998

[1]I thank audiences at numerous universities and conferences and Luca Anderlini, Jim Bergin, Eddie Dekel, Itzhak Gilboa, Ed Green, Joe Halpern, Debra Holt, Preston McAfee, Phillipe Mongin, Yaw Nyarko, Christine Parlour, Michele Piccione, Roy Radner, Ariel Rubinstein, Aldo Rustichini, Peter Streufert, Lenore Zuck, and Hyun Song Shin and two anonymous referees for helpful suggestions and conversations. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. A preliminary version of this paper was circulated under the title "Decision Theory with Impossible Possible Worlds." Finally, I thank Penn for an enjoyable visit during which some of this research was done.

Abstract

I propose modeling boundedly rational agents as agents who are not logically omniscient — that is, who do not know all logical or mathematical implications of what they know. I show how a subjective state space can be derived as part of a subjective expected utility representation of the agent's preferences. The representation exists under very weak conditions. The representation uses the familiar language of probability, utility, and states of the world in the hope that this makes this model of bounded rationality easier to use in applications.

1 Introduction and Motivation

This paper proposes a framework for the axiomatic development of a tractable model of bounded rationality.[1] As has long been argued, there are many economic phenomena which seem difficult to understand without a model of bounded rationality. Clearly, such a model must be tractable if it is to be useful in enlarging our understanding of economic phenomena. My approach is to develop a model of bounded rationality which uses the familiar notions of utility, probability, and states of the world in the hope that this makes the model easier to use in applications. While I do not provide applications here, the conclusion discusses some possibilities.

The key idea of my approach is to develop a decision theory which does not assume that agents are, in the phrase of some philosophers, *logically omniscient.* An agent is said to be logically omniscient if he knows all logical implications of his knowledge. It goes without saying that this is not a characteristic of real people. A real person can know the axioms of set theory and the rules of logical inference without knowing all the theorems of set theory, though these are logically implied.

Several kinds of bounded rationality can be seen as a lack of logical omniscience. First, the common criticism that real agents cannot compute the optimal action in a complex model is precisely a statement that real agents are not logically omniscient. To see this, consider the following fanciful example. Suppose we ask an agent to make a choice between \$100 and a box. We tell the agent, "The box contains $f(1)$ dollars where the function f is" and then we write the following on a convenient blackboard:

$$f(x) = \begin{cases} 0, & \text{if the } 10,000^{\text{th}} \text{ digit of } \sqrt{2} \text{ is } x; \\ 200, & \text{otherwise.} \end{cases}$$

For the purpose of this discussion, assume that $f(1) = 200$. If we believe that this person is rational according to our usual definition, what will we predict he will do? Since this person knows what the function f is — after all, it is written down in front of him — the usual economic analysis would conclude that a rational person chooses the box. The key issue, though, is not whether the person knows the function f but whether he knows its value at 1. If he knows its value at 1, we would certainly expect him to choose the box. But of course, it is difficult to imagine that a real person would happen to know the value of $f(1)$!

In other words, our usual approach takes logical omniscience as given. Clearly, given logical omniscience, the agent must deduce from the information on the blackboard that there is \$200 in the box. In a strictly logical sense, the description of the function f

[1] There are numerous models of bounded rationality in the literature, but few with an axiomatic basis. For surveys of the literature, see Lipman [1995] and Rubinstein [1998].

written down for the agent[2] is equivalent to a particular set of ordered pairs. Hence if one knows this description and knows all logical implications of what one knows, one must know the function f in the sense of knowing its value at 1 and at all other points in its domain. Of course, it is implausible that any real agent would immediately know all this information.

Dropping the logical omniscience assumption also allows us to study at least some kinds of *framing effects*. I use this term to refer to situations where the form of the information received by the agent has an effect on the decision. That is, two different but logically equivalent pieces of information lead to different choices. It is easy to see that if an agent is not logically omniscient, he may not view two "truly" equivalent pieces of information as equivalent and hence may react differently to the two.[3]

So far, I have been very informal about logical omniscience. In particular, the reader may well wonder what it is about the usual subjective expected utility model which presumes logical omniscience. I believe that the problem lies in the exogeneity of the set of states of the world.

The standard interpretation of a state of the world is as a complete, consistent description of a way the world might be. The logical omniscience assumption is precisely the consistency part of this definition. To see why, suppose that every state is, in a strictly logical sense, internally consistent. Suppose we define knowledge of a fact to mean that it is true in every state of the world — loosely, that the agent assigns it probability 1. Suppose then that the agent knows the axioms of set theory and the rules of logical inference, so that these hold in every state of the world. Then if all the states are consistent, this must mean that all the theorems of set theory are true in all states and so the agent must know these as well.

This is problematic because the traditional approach treats the options the agent has as *acts*, functions from states to consequences. In the example above, we would think of the option of the \$100 as the constant function at \$100 — that is, the function mapping every state into the consequence \$100. The option of the box would correspond to a function which gives zero dollars in any state in which $f(1) = 0$ and \$200 in any other state. The problem then is that if all the states are logically consistent, then the set of states in which $f(1) = 0$ is empty, so this act, in fact, gives \$200 in every state. In short, we are unable to distinguish the second choice from a sure receipt of \$200 and hence are unable to include in the model the uncertainty that seems so obviously present.

[2] Together with the definition of square roots, of course. The reader who is concerned about such issues is urged to assume that we also place in front of the agent a wide variety of mathematics texts which provide all the necessary definitions.

[3] Some seem to interpret framing effects as allowing situations where the agent may know the two forms of the information are equivalent but reacts differently anyway. I am not opposed to this view but it is not the kind of framing I consider.

The fact that dropping logical omniscience requires allowing inconsistent states is well–known in the philosophy literature. As noted by Hintikka [1975], among others, to use states of the world forces us to make one of two assumptions. Either we must assume that the agent knows all logical implications of his knowledge or we must assume that some of the states of the world are logically inconsistent.[4] Hintikka refers to such states as *impossible possible worlds* since they are logically impossible but are used to model what the agent considers possible.[5]

So far, I have only suggested that the usual interpretation of states is problematic if we wish to drop the assumption of logical omniscience. Why does this require anything more than a reinterpretation of the usual Savage framework? Why can't we continue to take the set of states to be exogenous but interpret them as complete though not necessarily consistent descriptions of how the world might be?

This approach runs into three serious conceptual problems. To explain the first, recall that in the standard subjective expected utility framework, the agent has preferences over all acts — that is, every function from the set of states to the set of consequences. If we enlarge the state space to its largest possible extreme, then we correspondingly enlarge the set of acts the agent must compare.[6] This enlargement can be very dramatic. To see this, consider the $\sqrt{2}$ example. Suppose the agent can bet on any of the first 10,000 digits and may think that any given digit could be any number. Then we will require at least 100,000 states, even though there is only one truly possible state of the world since the function f is nonstochastic. Hence if the agent can bet any amount of money, the set of acts changes from $\mathbf{R}$ to $\mathbf{R}^{100,000}$! It seems odd to study bounded rationality by assuming the agent's preferences satisfy the usual assumptions but making him compare a vastly larger set of alternatives.

To understand the second problem, change the $\sqrt{2}$ example above so that the definition of f is now that $f(x) = 200x$. Suppose the agent understands what multiplication is, but somehow doesn't realize that $f(1) = 200$. Then we have to assume that there is at least one state s such that the value of $f(1)$ at s is different from 200 — that is, at s, 200 times 1 is not equal to 200. Clearly, s is a very strange state of the world. What justification is there for including such peculiar states of the world in our model?[7]

[4] A similar observation is made by Rubinstein [1991; 1998, Chapter 4.7] in the context of imperfect recall. See also Fagin, Halpern, Moses, and Vardi [1995, Chapter 9].

[5] To prevent possible misunderstandings, I should mention that I do not use the phrase impossible possible worlds to refer to the specific formulation of this idea proposed by Rantala [1975].

[6] Of course, one may be able to obtain an expected utility representation with only a subset of the possible acts. However, the larger the set of states, the larger the set of acts we will need the agent to compare.

[7] It bears emphasizing that the problem with s is not simply that it is the "wrong" state of the world in the sense that it does not correspond to reality. If, in fact, it rains today, those states where it does not rain do not correspond to reality. The distinction is that we can construct *internally* consistent circumstances in which it does not rain today, though these circumstances are inconsistent with *external*

Finally, and perhaps most importantly, the usual approach to subjective expected utility relies heavily on the properties of conditional preferences. Loosely, Savage requires that preferences conditional on learning any given event in the state space are well behaved. To see why this assumption becomes highly questionable if we include inconsistent states, suppose we have an agent choosing between three envelopes, labelled a, b, and c, each containing an unknown amount of money. Do we really expect the agent to have sensible preferences conditional on learning that envelope a has \$10 more than b, b has \$10 more than c, and c has \$10 more than a?

Perhaps for these reasons, Savage himself was quite skeptical about extending subjective expected utility theory to eliminate the logical omniscience assumption. In 1967, he wrote:

> [T]he postulates of the theory [of subjective expected utility] imply that you should behave in accordance with the logical implications of all that you know. Is it possible to improve the theory in this respect ... or would that entail paradox, as I am inclined to believe but am unable to demonstrate?

The framework I propose endogenizes the state space by making the impossible worlds part of our representation of the agent's preferences, just like his utility function or subjective probabilities. This approach avoids all three of the difficulties noted above with simply enlarging the state set to the largest possible extreme. First, the set of acts is not enlarged in my approach because the enlargement of the state space is part of our representation, not part of the primitives. I start with the set of truly possible states and the usual set of acts. I then derive the impossible possible worlds as part of a representation of preferences over these acts and hence never need to enlarge the set of acts.

Second, my approach gives a simple answer to the natural question: what impossible worlds do we add and why? In a sense I make precise later, I use the agent's preferences to identify the relationships between pieces of information he perceives and then construct his state space to reflect this. For instance, in the $\sqrt{2}$ example above, if the agent prefers the envelope, this preference tells us that he believes it possible that $f(1) = 0$. Hence our representation must include an impossible world in which this is true.

Finally, while conditional preferences are an important part of the construction, I only consider preferences conditional on *consistent* information. Because the impossible worlds are only part of the representation, preferences conditional on such worlds cannot be relevant.

facts. However, we *cannot* construct internally consistent circumstances in which multiplication works as usual and yet 200 times 1 is not 200.

In addition, the axiomatic approach also has several advantages. First, it provides a way to nest a variety of different models within a single framework, thus providing a tool for the comparison of models which are formulated very differently.

Second, the axiomatic approach is immune to some criticisms of models of bounded rationality. One criticism is that these models seem *ad hoc*. I believe this often means no more than the fact that the models are not standard. However, it is certainly true that we rarely have a clear reason for adopting one model of bounded rationality over another. In the axiomatic approach, it is quite clear what justifies the model: it is a simple way to represent the agent's preferences when those preferences satisfy certain conditions. In this sense, the impossible worlds are no more *ad hoc* than utility functions or subjective probabilities.

A second criticism concerns the fact that most models of bounded rationality assume that the agent deals with his limitations optimally. Many people object to such a treatment, arguing that we have assumed that the agent is able to solve a more difficult problem than the one we began by assuming he could not solve! Of course, extending the model to allow boundedly rational choice of computations seems to lead us into an infinite regress problem.[8] Again, this point is moot when the model is a representation of preferences. It makes no sense to ask the question "how can the agent carry out this complex task?" when the "complex task" is simply our representation of whatever it is the agent in fact manages to do.

The rest of this paper is organized as follows. In Section 2, I give the basic framework for analysis and present the main results. In particular, I give necessary and sufficient conditions for deriving a subjective state space from the agent's preferences as well as necessary and sufficient conditions for representing the agent's preferences via expected utility on this state space. In Section 3, I give two examples illustrating the ideas. In the examples, simple assumptions on the agent's reasoning translate into remarkably clean restrictions on the way the impossible worlds are constructed. Section 4 offers a few concluding remarks on possible applications and extensions. All proofs are contained in the Appendix.

Related Literature. As far as I know, there has essentially no formal consideration of the logical omniscience problem in the economics literature. Perhaps the closest predecessor is Shin's [1993] demonstration that if knowledge is interpreted as provability, then agents will not generally satisfy the negative introspection property.

On the other hand, the idea of enlarging the state space in order to represent something nonstandard on the original state space is an old one. However, there has been relatively little work deriving an enlargement from preferences. Kreps [1979, 1992] showed

[8] See Lipman [1991] for further discussion and a different approach to resolving this problem.

that a preference for "flexibility," interpreted as a recognition by the agent that not all possibilities are foreseen, can be represented via a particular extension of the state set. As I discuss later (see Remark 4), there are some unexpected technical similarities between Kreps' analysis and mine. Gilboa and Schmeidler [1994] show that Choquet expected utility is equivalent to expected utility on an enlarged state space. A key aspect of their results is the well–known fact that belief functions — nonadditive functions on a state set — can be derived from additive functions on a larger state set (see, for example, Shafer [1976]). See the discussion following Theorem 5 for more details on this connection.

Another related paper is Morris [1996]. He assumes logical omniscience but uses preferences to derive a belief operator for the agent (a nonpartitional information structure) in a manner akin to my use of preferences to derive the agent's subjective state space. In effect, we both use the way preferences vary with information to identify a model of information processing.

Finally, there have been many models without logical omniscience proposed in the philosophy and artifical intelligence literature. For a survey, see Fagin, Halpern, Moses, and Vardi [1995]. Some concrete connections to this literature are discussed in Lipman [1994a, 1994b].

2 The Model

Notational Conventions. For any sets A and B, A^B denotes the set of all functions $f : B \to A$ and 2^A the set of all subsets of A. If B is a collection of sets, then $\cap B$ is the intersection of all the sets in B.

The usual approach treats information as the ruling out of states. Logically equivalent pieces of information would rule out the same set of states, so this approach does not allow one to distinguish between different statements which are logically equivalent. Since I require such a distinction, I take a more abstract approach, treating pieces of information simply as points in a set. Let Φ denote the set of pieces of information. Elements of this set can be thought of as propositions (more precisely, propositional formulae) in logic, statements in English or another language, or mathematical formulas.

I define "correct" logical deduction by means of a nonempty collection of subsets of Φ, denoted $\mathcal{I}$. The sets in this collection are internally consistent sets of information — that is, a set is defined to be internally consistent if and only if it is in the collection $\mathcal{I}$. I refer to an element I of $\mathcal{I}$ as an *information set* to emphasize the interpretation that I is one possible set of information the agent might receive.

I now use $\mathcal{I}$ to define the "truly possible" states of the world. To do so, recall the usual informal definition of a state as a complete and logically consistent description of how the world might be. To model this, I will treat states as subsets of Φ, with the interpretation that at state $s \subseteq \Phi$, any piece of information in the set s is true and anything not in s is false. Since states are supposed to be logically consistent, I require states to be elements of $\mathcal{I}$. Since they are supposed to be complete, I require them to be maximal (under $\subseteq$) elements. In short, S, the set of *possible worlds*, is defined to be the set of maximal elements of $\mathcal{I}$. I reserve s to denote a typical element of S, though s', s^*, etc., may be another kind of subset of Φ, such as the impossible worlds.

For any $I \subseteq \Phi$, let

$$S(I) = \{s \in S \mid I \subseteq s\}.$$

That is, $S(I)$ is the collection of states of the world in which each $\varphi \in I$ is true or, more briefly, in which I is true. If $S(\{\varphi\}) = \emptyset$, I will say φ is a *contradiction*.

Throughout, I make the following assumptions on $\mathcal{I}$:

Assumption 1

(1) $$I \in \mathcal{I},\quad I' \subseteq I \implies I' \in \mathcal{I}$$

(2) $$I \in \mathcal{I} \implies S(I) \neq \emptyset$$

(3) $$S \text{ is finite}$$

Condition (1) is the natural requirement that any subset of a consistent set is itself consistent. (2) says that the set of information sets is closed in the sense that it requires the limit of any increasing sequence of information sets to be an information set.[9] Finally, (3) is useful for simplifying the analysis.

Example. Suppose that $\Phi = \{p, \neg p, p \text{ or } \neg p\}$ where $\neg p$ is interpretted as the negation of p. It is natural to define a subset of Φ to be consistent if it does not include *both* p and $\neg p$. If we adopt this definition, then

$$\mathcal{I} = \{\emptyset, \{p\}, \{\neg p\}, \{p \text{ or } \neg p\}, \{p,\quad p \text{ or } \neg p\}, \{\neg p,\quad p \text{ or } \neg p\}\}$$

It is easy to see that there are two states, $\{p,\quad p \text{ or } \neg p\}$ and $\{\neg p,\quad p \text{ or } \neg p\}$. Clearly, Assumption 1 is satisfied. ▮

Note that, as in the example, condition (1) of Assumption 1 implies that $\emptyset \in \mathcal{I}$. I view the case where the agent's information set is $\emptyset$ as the *ex ante* situation, where the agent has received no information.

[9] For an example of what is being ruled out, suppose that Φ is the integers and that $\mathcal{I}$ is the collection of all finite subsets of Φ. Then we have $S = \emptyset$, violating (2).

Remark 1 One example satisfying Assumption 1 is propositional logic. Suppose that Φ is the set of all propositional formulae generated from some set of atomic formulae. Suppose we define $\mathcal{I}$ and S using the usual propositional logic. It is easy to see that conditions (1) and (2) of Assumption 1 would automatically be satisfied and (3) would hold if we assume that the set of atomic formulae is finite.

Let X be the set of *consequences*. The interpretation of a consequence, as in Savage [1954], is that it is a sufficiently detailed description of the outcome of a choice to determine the way the agent evaluates that outcome. For simplicity, I will take $X = \mathbf{R}$. Let $F = X^S$ denote the set of *acts*. Again as in Savage [1954], actions are treated as simply creating a relationship between states and consequences. In the conclusion, I briefly discuss an alternative formulation of acts which I plan to pursue in future work.

The key ingredient of the framework is a collection of binary relations on F, one for each information set. I let $\succ_I$ denote the agent's preferences given information set I and $\{\succ_I\}$ the collection of these preferences. As usual, $f \sim_I g$ denotes $f \not\succ_I g$ and $g \not\succ_I f$ while $f \succeq_I g$ if $f \succ_I g$ or $f \sim_I g$.

It is important to clarify the interpretation of $\{\succ_I\}$. First, $\succ_I$ is the agent's preference *after* whatever processing of I he carries out. That is, I assume the agent receives information set I, analyzes this information in whatever fashion he chooses, and then he reaches some perception of the problem. $\succ_I$ is intended to represent this final perception of the problem — the perception he acts on.

Second, 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 information set, not that the agent himself knows this *ex ante*.

Third, as discussed in the introduction, preferences are only defined at information sets, so I do not consider the agent's preferences in response to nonsensical information.

Finally, it bears emphasizing how and why this approach differs from the traditional one. Normally, the primitive of the model is a single *ex ante* or "informationless" preference relation. From this, one derives preferences conditional on information under some assumptions on how the agent responds to information. Here the various preferences are the primitives and it is a characterization of the relationship between them that is to be derived. This approach is followed because the goal of the paper is to uncover a representation of the agent's information processing from these different preference relations, rather than the more usual reverse procedure.[10] Using weaker assumptions does

[10]Luce and Krantz [1971] provided the first explicit use of a family of conditional preferences in place of a single preference relation. Since their motivation is different from mine, they use preferences conditional on an event in the state space and do not construct a state space. For a more recent example

not come without a cost: it is not clear whether one can observe a large enough set of an agent's different conditional preferences to be able to provide a conclusive test of the model.

For simplicity, I assume the following throughout the paper.

Assumption 2 *There are finitely many distinct preference relations in* $\{\succ_I\}$.

A natural way to try to represent these preferences would be with standard expected utility. To state this more precisely, I will say that an information set I is *null* if $f \sim_I g$ for all $f, g \in F$.

Definition 1 $\{\succ_I\}$ *is* expected utility (EU) representable *if there is a function* $u : X \to \mathbf{R}$ *and a probability measure* μ *on* S *such that for all nonnull* I, $\mu(S(I)) > 0$ *and*

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

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

It is straightforward to restate the Savage [1954] axioms in this framework to give sufficient conditions for such a representation.[11]

There is one necessary condition for an expected utility representation which is implicit in the usual framework and so is not normally discussed.

Definition 2 *Information sets* I *and* I' *are* logically equivalent *if* $S(I) = S(I')$.

Definition 3 $\{\succ_I\}$ respects logical equivalence *if for every logically equivalent* I *and* I', *we have* $\succ_I \; = \; \succ_{I'}$.

Proposition 1 *If* $\{\succ_I\}$ *is EU representable, then it respects logical equivalence.*

---

of this approach, see Skiadis [1996].

[11] The finiteness of the state space does complicate matters. See Gul [1992] and Chew and Karni [1994].

The proof of this result is obvious: if I and I' are logically equivalent, then the updating in response to I must be the same as the updating in response to I'. That is, $S(I) = S(I')$, so the expected utility of any act conditional on I must equal the expected utility of the same act conditional on I'. Hence the preferences must be the same. In this sense, the usual expected utility approach requires that the agent recognize all logical equivalences. Putting it differently, framing effects are inconsistent with expected utility.

It is important to note the source of the difficulty: it is not that utility is represented by an additive function across states nor the particular use of Bayes' Rule to calculate updated probabilities. Instead, it is a consequence of treating information in terms of the event of the state space it logically corresponds to.

My approach is based on this observation: to deal with the logical equivalences the agent doesn't recognize, I identify the equivalences he does recognize and construct a state set in which these are the "correct" equivalences. That is, if I and I' are equivalent but framing effects lead the agent to react differently to the two, then add states — impossible possible worlds — to his subjective state space in which one information set is true and the other is not.

To be more precise, given any $\{\succ_I\}$, we can define an equivalence relation $\cong$ on $\mathcal{I}$ by

$$I \cong I' \iff S(I) = S(I') \text{ and } \succ_I \;=\; \succ_{I'}.$$

That is, $I \cong I'$ means that I and I' are logically equivalent and the agent responds identically to these two information sets. Hence we can treat the agent as correctly recognizing this logical equivalence. When $I \cong I'$, I say that I is *recognized equivalent* to I'. Loosely, $I \cong I'$ means that the difference between I and I' does not create a framing effect.

The first step is to extend the state set in a way which captures equivalence according to $\cong$. Given some collection of sets containing pieces of information, $S^* \subseteq 2^\Phi$, let

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

Definition 4 $S^* \subseteq 2^\Phi$ preserves $\cong$ *if* $S \subseteq S^*$ *and*

$$I \cong I' \iff S^*(I) = S^*(I').$$

In other words, S^* preserves $\cong$ if the "logical" equivalence of information sets in S^* corresponds exactly to $\cong$. In addition, I impose the requirement that $S^* \supseteq S$. Without this, one could discard some or all of the original states, which seems unreasonable to me. I believe that this requirement is not important to the results, however.

If the agent does not recognize all logical equivalences, we cannot represent him via expected utility on S. However, S^* is constructed precisely so that "logical" equivalences on S^* are the equivalences he recognizes. Hence we may be able to represent him via expected utility on S^*. The fact that we have added states means that we must specify the consequences of the acts on these new states. This suggests the following approach.

Definition 5 $\{\succ_I\}$ *is* extended expected utility (Extended EU) representable *if there exists an* extended state set $S^* \supseteq S$, $S^* \subseteq 2^\Phi$, *an* act extension $h : F \to X^{S^*}$, *a utility function* $u : X \to \mathbf{R}$, *and a probability measure* μ *on* S^* *such that* S^* *preserves* $\cong$,

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

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

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

Remark 2 An Extended EU representation requires that $\mu(S(I)) > 0$ if I is nonnull. This is stronger than the perhaps more natural requirement that $\mu(S^*(I)) > 0$ (note that $S(I) \subseteq S^*(I)$). However, without this requirement, Extended EU representation becomes almost trivial since one can choose a μ such that $\mu(S) = 0$.

An Extended EU representation must (a) preserve the logical equivalences the agent recognizes and (b) represent his preferences in an analogous fashion to the usual expected utility representation. Each of these requirements translates into restrictions on the preferences which can be represented this way.

One obvious restriction imposed by the second requirement is

Definition 6 $\{\succ_I\}$ *is* representable *if for all nonnull* I, *there exists* $u_I : F \to \mathbf{R}$ *such that*

$$f \succ_I g \iff u_I(f) > u_I(g).$$

Clearly, representability is necessary for $\{\succ_I\}$ to be Extended EU representable — without it, preferences conditional on some nonnull I are not representable by a utility function at all, much less one with the particular structure Extended EU requires. Necessary and sufficient conditions for representability are well known so I will omit discussion of them.

Remark 3 While representability is a weak assumption, I do not wish to claim that it is innocuous for the study of boundedly rational agents. However, there are two reasons for not insisting on a weaker assumption at this point. First, as suggested in the introduction, one important aspect of bounded rationality does seem to be the fact that real agents are not logically omniscient. In a sense, assuming representability says that I am isolating this aspect of bounded rationality for study rather than trying to capture all of what bounded rationality means in a single model. Second, I believe that representability can be replaced with weaker assumptions if we replace Extended EU with a weaker form of preference representation. For example, we might replace u_I with some representation which allows for intransitive indifference or other limitations on perceptual ability (see Fishburn [1973] for a survey) and obtain a generalization of Extended EU.

Next, I wish to develop the restrictions imposed by the requirement that S^* preserve $\cong$. These restrictions are most easily stated in terms of an ordering on $\mathcal{I}$ which can be interpreted as "more informative than." Intuitively, the subjective state space we are constructing is supposed to reflect the agent's perception of the relationships between different pieces of information. I have already pinned down the connection partly by requiring that when $I \cong I'$, then I and I' induce the same event in the agent's state space. What is missing is a way to relate events for nonequivalent information sets. Of course, unrelated information sets will not induce events which are nicely related. So what relationship between information sets should we focus on?

The key is when one information set conveys more information to the agent than another. We would naturally expect that if one information set is more informative to the agent than another, then it pins down the state more precisely in the sense that it induces a smaller event in the agent's state space. In other words, if I is more informative than I', denoted $I \succeq^* I'$, then $S^*(I) \subseteq S^*(I')$.

How would we identify $\succeq^*$? Intuitively, if $I \cong I'$, then each information set is weakly more informative than the other. Also, suppose that adding I' to I does not add any information to that conveyed by I alone. That is, $I \cong I \cup I'$. Then it seems sensible to say that I conveys more information than I'. In line with this intuition, I define $\succeq^*$ as follows.

Definition 7 *Given a recognized equivalence relation $\cong$, the associated* information ordering *$\succeq^*$ is defined by*

$$I \succeq^* I' \iff I \cong I \cup I' \text{ or } I \cong I'.$$

Remark 4 The $\succeq^*$ relation is similar to the domination relation studied in Kreps [1979]. He considers preferences over subsets of some set of alternatives and says that set x

dominates x' if x is indifferent to $x \cup x'$, analogously to the way I is recognized equivalent to $I \cup I'$. The construction of states he gives is similar to mine, though his use of these states and the interpretation of the order generating them are quite different.

The following theorem indicates that this definition of the information ordering does appropriately identify the agent's view of "more informative than."

Theorem 1 *S^* preserves $\cong$ if and only if*

$$I \succeq^* I' \iff S^*(I) \subseteq S^*(I').$$

In other words, if there is a extended state set which appropriately represents $\cong$, then the information ordering will do what it is supposed to: it tells us when $S^*(I) \subseteq S^*(I')$.

The agent's information ordering gives a very simple way to state the conditions which enable us to construct an appropriate extended state set. The crucial condition is:

Definition 8 $\{\succ_I\}$ *is* consistent *if* $\succeq^*$ *is transitive and*

$$I \succeq^* I' \iff I \succeq^* \{\varphi\}, \ \forall \varphi \in I'. \tag{1}$$

For brevity, I often say that $\cong$ or $\succeq^*$ is consistent in place of the more precise statement that the underlying $\{\succ_I\}$ is consistent.

These properties seem not unreasonable for a subjective notion of "more informative than," though it is important to note that they do rule out some interesting forms of bounded rationality. While transitivity seems a fairly natural requirement, it does rule out resource–bounded reasoning, where an agent has a fixed amount of time (or other resources) to use for computation. To see this, suppose that if the agent learns I, he can deduce I' but no more, while if he began with the information I', he could deduce I''. It seems quite reasonable to believe that if told only I, he could not deduce I''. In this case, we would expect to have $I \succeq^* I'$ (since I informs the agent of I'), $I' \succeq^* I''$, but not $I \succeq^* I''$, a violation of transitivity. Such intransitivities are ruled out by the existence of a subjective state space because $\succeq^*$ is represented by a subset relation on the subjective state space and, of course, this relation is transitive.

As for (1), one direction seems very compelling: if I is more informative than I', surely this means that I is more informative than any single element of I'. The converse is not implausible but is certainly not as compelling. This assumption rules out the quite plausible situation where the combination of two pieces of information is more informative than the pieces separately — where, for example, I could be more informative than $\{\varphi\}$ and more informative than $\{\psi\}$ but less informative than the two together. To see why this possibility is precluded by the subjective state space, note that if I is more informative than each of $\{\varphi\}$ and $\{\psi\}$, then $S^*(I) \subseteq S^*(\{\varphi\})$ and $S^*(I) \subseteq S^*(\{\psi\})$. But then

$$S^*(I) \subseteq S^*(\{\varphi\}) \ \cap \ S^*(\{\psi\}) = \{s^* \in S^* \mid \varphi \in s^* \text{ and } \psi \in s^*\} = S^*(\{\varphi, \psi\}).$$

Another way to get some intuition for consistency is by examples. It is not hard to show that if the agent's preferences respect logical equivalence (that is, $I \cong I'$ whenever $S(I) = S(I')$), then they are consistent. At the opposite extreme, if the agent recognizes no equivalences (other than recognizing that I is equivalent to itself), his preferences again satisfy consistency.

The next theorem shows that consistency is what is needed to construct the subjective state space.

Theorem 2 *There is an $S^* \supseteq S$ which preserves $\cong$ if and only if $\{\succ_I\}$ is consistent.*

In short, for $\{\succ_I\}$ to be Extended EU representable, it must be representable and consistent. Theorem 3 states that these two conditions plus either of two additional assumptions are sufficient for Extended EU. Neither additional assumption is necessary, nor is it necessary that at least one of the two holds. The first of the additional conditions is:

Definition 9 Φ *is* broad *if it contains at least one φ such that $S(\{\varphi\}) = \emptyset$.*

In other words, Φ is broad if there is at least one contradiction in Φ. This condition seems quite innocuous but is less economic than the following alternative:

Definition 10 $\{\succ_I\}$ *satisfies* weak state independence (WSI) *if there exists an onto function $u : X \to \mathbf{R}$ such that whenever $I \cong s$ and I is nonnull, we have*

$$f \succ_I g \iff u(f(s)) > u(g(s)).$$

The primary requirement of WSI — that preferences conditional on certain information sets depend only on the consequence associated with the identified state — is weaker than Savage's state independence condition, P3. Savage's P3 is essentially an ordinal version of WSI but applied to every information set for which the act is constant on the associated set of states, rather than just those information sets which are recognized equivalent to a state. WSI also has a cardinal side to it, requiring these preferences to be representable with a utility function. The existence of a function u is implied by representability and so is not an additional restriction. The requirement that u is onto is an additional restriction, essentially requiring there to be no best or worst consequences and no "gaps."

Theorem 3 *If* $\{\succ_I\}$ *is Extended EU representable, then it is representable and consistent. If* $\{\succ_I\}$ *is representable and consistent and* either Φ *is broad or* $\{\succ_I\}$ *satisfies weak state independence, then it is Extended EU representable.*

3 Examples

Recapping, the agent's preferences $\{\succ_I\}$ naturally identify an information ordering we can attribute to him. If his preferences are consistent, we can use this to identify an extended state set and represent the agent as an expected utility maximizer on this state space. Unfortunately, the extended state set is not unique. (Appendix D gives a characterization of the extended state set showing the extent to which it is identified.) As the following examples show, however, one can use simple and not implausible restrictions on the agent's reasoning ability to generate useful restrictions on the agent's subjective state space.

To state the restrictions most simply, I make the following assumption throughout this section:

Assumption 3 Φ *is* not *broad.*

In other words, I assume throughout this section that Φ contains no contradictions. Remark 6 explains how to modify the statements of the results for the case where Φ is broad.

3.1 Example 1

One not implausible property for $\succeq^*$ to satisfy is

Definition 11 $\{\succ_I\}$ respects simple implication *if* $S(\{\varphi\}) \subseteq S(\{\psi\})$ *implies* $\{\varphi\} \succeq^* \{\psi\}$.

Intuitively, if φ implies ψ in standard logic, then the agent recognizes that φ tells him that ψ holds, so φ conveys at least as much information to him as ψ. The phrase "simple implication" is meant to focus on the fact that the condition only applies when the premise and the conclusion each consist of a single statement.

This property has remarkably clean implications for the determination of the agent's subjective state space. To explain this, let $S^*_\cap$ be the collection of sets we can construct by taking intersections of possible worlds. More formally,

$$S^*_\cap = \{s^* \subseteq \Phi \mid s^* = \cap B, \text{ for some } B \subseteq S\}.$$

Let τ denote the smallest topology on Φ containing S. The finiteness of S implies that τ is the topology generated by the base $S^*_\cap$. More precisely, τ is simply the collection of sets generated by taking unions of sets in $S^*_\cap$. In short, $s^* \in \tau$ if and only if there is a collection $\mathcal{B}$ of subsets of S such that

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

Theorem 4 *If S^* preserves $\cong$, then $S^* \subseteq \tau$ if and only if $\{\succ_I\}$ preserves simple implication.*

In other words, given an intuitive, though not trivial, restriction on the information ordering, we can restrict attention to a very simple procedure for constructing impossible worlds — namely, we only need to consider sets formed by taking unions of intersections of truly possible worlds. The following example gives a simpler procedure, though for a more restrictive case.

3.2 Example 2

Recall that $\{\succ_I\}$ respects logical equivalence if $S(I) = S(I')$ implies $\succ_I \;=\; \succ_{I'}$. Clearly, this is too strong a property for boundedly rational agents. However, it is of interest for two reasons. First, it provides a particularly clean illustration of the construction of S^*. Second, as we will see, it provides new insights into decision theory for unboundedly rational agents.

The characterization of S^* for this case is particularly nice when the "language" is sufficiently expressive. The specific notion is:

Definition 12 Φ *is* rich *if for every nonempty* $\hat{S} \subseteq S$*, there exists* $\varphi \in \Phi$ *such that* $S(\{\varphi\}) = \hat{S}$*.*

For example, the usual propositional logic (when S is finite) satisfies this condition.

Theorem 5 *If* Φ *is rich and* S^* *preserves* $\cong$*, then* $S^* \subseteq S^*_\cap \cup \{\Phi\}$ *if and only if* $\{\succ_I\}$ *preserves logical equivalence.*

Combined with Theorem 3, this result has a surprising implication. Suppose we focus on the case of a "perfectly rational" agent — that is, one who recognizes all logical equivalences. Suppose we make only minimal regularity assumptions on the preferences — namely, Assumption 2, representability, and weak state independence. Note, in particular, that no sure–thing principle assumption is made at all. Then we can extend the state set to $S^*_\cap \cup \{\Phi\}$ and represent these preferences by expected utility. It is not hard to show that one doesn't need to include a state equal to Φ,[12] so using $S^*_\cap$ as a state set is sufficient. In short, the extension of the state set, aside from capturing imperfect reasoning, can also "rectify" failures of the sure–thing principle.

This result is a generalization of Gilboa and Schmeidler [1994]. They show that Choquet expected utility with capacities that are belief functions is equivalent to standard expected utility on an enlarged state set. While their framework is different, it is not difficult to show that the appropriate analogue of their enlarged state set is $S^*_\cap$.[13] In a related vein, when Shafer [1976] shows that belief functions are equivalent to probability distributions on an enlarged set, the enlarged set is again the analogue in his framework of $S^*_\cap$. Theorem 5 indicates why it is $S^*_\cap$ which appears in these contexts — it is precisely the way to extend the state space when agents reason correctly, an assumption maintained in both of these works.

Remark 5 $S^*_\cap$ also appears in the work of Rescher and Brandom [1979], where this is one of the two types of impossible worlds they introduce to study inconsistency. The other type introduced involves taking unions of elements of S. It is easy to see that τ provides a natural generalization of these two constructions. See Lipman [1994a] for a more detailed discussion of this connection.

[12]This statement relies on weak state independence. If one assumes that Φ is broad instead of weak state independence, then it is necessary to include Φ as a state.

[13]To be more precise, the conclusion that the enlarged state set they consider is $S^*_\cap$ relies on the approach they take to updating as well. Interestingly, Ghirardato and Le Breton [1997] have recently shown that Choquet expected utility with general capacities is also equivalent to standard expected utility on an enlarged state set. Since they do not consider issues of updating, it is not completely clear what the analogous result here is, but the enlargement has a similar structure to τ.

Remark 6 It is simple to adapt the statements of Theorems 4 and 5 to the case where Assumption 3 is dropped, so Φ does contain some contradictions. Only two changes are required. First, in Theorem 5, $\{\Phi\}$ is replaced by $\{\bar{\Phi}\}$ where $\bar{\Phi}$ is the set of elements of Φ which are not contradictions. To understand the second change, note that since contradictions cannot appear in any information set, they are irrelevant to whether a given state set preserves the agent's recognized equivalences. In light of this, we can restate the theorems above as saying that the given property of $\{\succ_I\}$ holds if and only if after removing all contradictions from all states, the resulting state set satisfies its given property.

4 Conclusion

In this paper, I have proposed a framework for an axiomatic study of bounded rationality. This approach constructs a state set which represents the agent's view of how pieces of information are related to one another. I then represent the agent's preferences using expected utility on this state set. Because the approach uses familiar notions like probability, utility, and states of the world, my hope is that it may prove more useful in applications than models of bounded rationality based on less familiar concepts.

One very natural direction for application is framing effects. The framework crucially relies on the idea that the agent may react differently to two equivalent pieces of information, precisely what framing effects are all about. Hence this approach may prove useful for more detailed models of framing effects, when they are likely to occur, how they affect decisionmaking, etc. For example, one might use this model to study advertising or other forms of persuasion as the attempt to manipulate framing effects.

The framework also uses "language" (in the form of Φ) and the agent's interpretation of it in a critical way. This aspect of the model could be useful in a variety of applications. One particularly prominent example is contracts. In this context, the elements of Φ would include statements regarding the uncertainty the agents are contracting over. A contract, then, would be a function from Φ (or subsets of it) to outcomes or agreements. "Perfectly rational" agents would view S as the relevant state space and any proposed contract as the implied function from states into outcomes. Boundedly rational agents, however, would not necessarily translate the contract in this fashion, allowing, for example, the possibility of disagreement regarding what the contract means.

For a simple illustration of why this formulation of bounded rationality could have interesting effects, suppose that the contract is purely for risk–sharing purposes — say, one agent is risk averse and the other risk neutral. Then if the agents are perfectly rational, we would predict that the risk neutral agent would perfectly insure the risk averse one. Hence the contract would call for state–contingent trades which yield a consumption level for the risk averse agent which is constant over S. However, a boundedly rational agent would not necessarily view such a contract as involving perfect insurance — that is, consumption may not be constant over S^*. In the example in the introduction, the "bet" considered is nonstochastic and hence has no risk whatsoever for a perfectly rational agent. However, most of us would be unsure about the outcome! Hence boundedly rational agents will view contracts differently, possibly preferring contracts which are "simpler" or incomplete. Similarly, whether a given set of markets is complete or not will depend on the subjective state space of the agents. In particular, markets which are complete in an objective sense may be incomplete from the agents' point of view and so may behave more like incomplete markets.

There are also many interesting extensions of the model to consider. First, the extension function h is unrestricted here. This function is crucial in that it describes how the agent views the outcomes of the possible acts and so it is worth considering some natural restrictions on h. One possible approach to this problem is to generalize the notion of an act to be a function from Φ (or a subset of it) to consequences, analogously to the representation of contracts suggested above. In this way, we could drop the usual assumption that the agent perceives acts in the form of functions from a state space to consequences.[14]

A second extension of interest is to generate computation and computation costs as part of a representation of preferences. To see how this could be done, suppose we split information sets into two components, a "pure" set of information (relating to external events) and a statement of what feasible set of actions the agent will be choosing from. Holding the second component of the agent's information fixed, we can carry out the analysis above to identify a subjective state set for the agent. The way this state set varies with the feasible set can be used to uncover a notion of computation cost. Intuitively, we would expect the agent to do more computation and hence have a more refined state space if the "value of information" is higher. Since the feasible set determines this value, we can use variations in the agent's state set across feasible sets to identify the computation costs.

Finally, an important topic for applications is the extension to many agents. In the many agent setting, it is natural to consider what agent 1 deduces when he learns that agent 2 knows some given fact. In other words, one could consider information sets which include modal propositions like $k_2\varphi$ ("agent 2 knows φ"). In such a context, one could analyze the conditions under which agents could be represented via partitions of some commonly known state space. Such a result might generalize Mukerji and Shin [1997], in much the same way that Theorem 5 generalizes Gilboa and Schmeidler [1994].

[14] Kreps [1992] and Skiadis [1996] provide different approaches to deriving a subjective notion of acts. See Dekel, Lipman, and Rustichini [1997] for discussion.

A Preliminaries

Throughout the appendix, for any set A, $\#A$ denotes the cardinality of A. For a set of sets, say B, $\cup B$ is the union of all the sets in B. Also, for any information set I, let

$$\pi(I) = \bigcup\{I' \in \mathcal{I} \mid I' \cong I\}$$

and let

$$\Pi = \{\pi \mid \pi = \pi(I), \text{ for some } I \in \mathcal{I}\}.$$

The following lemma establishes some useful facts. (The mneumonic is that equations beginning with D give properties following from the definition, while those beginning with C follow from consistency.)

Lemma 1 *For any $\{\succ_I\}$, the associated $\cong$, π, and $\succeq^*$ satisfy:*

(D1) $$I \cong I' \iff I \succeq^* I' \succeq^* I$$

(D2) $$I' \subseteq I \implies I \succeq^* I'$$

(D3) $$\pi(I) \in \mathcal{I}, \quad \forall I \in \mathcal{I}$$

If $\{\succ_I\}$ is consistent, then the associated $\cong$, π, and $\succeq^*$ satisfy

(C1) $$I \succeq^* I', \ \forall I' \in \hat{\mathcal{I}} \subseteq \mathcal{I} \implies I \succeq^* \cup\hat{\mathcal{I}}$$

(C2) $$I \cong I', \ \forall I' \in \hat{\mathcal{I}} \subseteq \mathcal{I} \implies I \cong \cup\hat{\mathcal{I}}$$

(C3) $$I \subseteq \pi(I') \iff \pi(I) \subseteq \pi(I')$$

(C4) $$\pi(I) = \bigcap\{\pi' \in \Pi \mid I \subseteq \pi'\}$$

(C5) $$S \subseteq \Pi$$

(C6) $$\pi(\pi(I)) = \pi(I)$$

(C7) $$I \succeq^* I' \iff I \cong I \cup I'$$

Proof of Lemma. For (D1), note that $I \cong I'$ obviously implies $I \succeq^* I'$ and the reverse. For the converse, suppose $I \succeq^* I' \succeq^* I$. If either direction of $\succeq^*$ comes from $I \cong I'$, we are done. So suppose that $I \succeq^* I'$ comes from $I \cong I \cup I'$ and $I' \succeq^* I$ comes from $I' \cong I \cup I'$. Since $\cong$ is an equivalence relation, transitivity gives $I \cong I'$.

For (D2), note that $I' \subseteq I$ implies $I = I \cup I'$. Since $\cong$ is reflexive, then, we obtain $I \cong I \cup I'$, so $I \succeq^* I'$.

To show (D3), let

$$\hat{I} = \{I' \in \mathcal{I} \mid I' \cong I\}$$

so $\pi(I) = \cup\hat{I}$. Then

$$S(\pi(I)) = \bigcap_{I' \in \hat{I}} S(I').$$

But $I' \cong I$ implies $S(I') = S(I)$, so $S(\pi(I)) = S(I)$. Hence if I is an information set and so, by condition (2) of Assumption 1, has $S(I) \neq \emptyset$, then $S(\pi(I)) \neq \emptyset$. Hence there is an information set — specifically, any $s \in S(\pi(I))$ — which contains $\pi(I)$. Hence by condition (1) of Assumption 1, $\pi(I)$ is an information set as well.

Remark 7 (D3) implies that I can consider whether $\pi(I)$ is recognized equivalent to some other information set. I will often use (D3) for this purpose without comment.

To show (C1), suppose $I \succeq^* I'$ for all $I' \in \hat{\mathcal{I}}$. By consistency,

$$I \succeq^* \{\varphi\}, \quad \forall \varphi \in I', \ \forall I' \in \hat{\mathcal{I}}$$

implying $I \succeq^* \cup\hat{\mathcal{I}}$.

For (C2), first note that (D1) implies

$$I \cong I', \ \forall I' \in \hat{\mathcal{I}} \iff I \succeq^* I' \succeq^* I, \ \forall I' \in \hat{\mathcal{I}}.$$

By (C1), then, $I \succeq^* \cup\hat{\mathcal{I}}$. By (D2), $\cup\hat{\mathcal{I}} \succeq^* I'$ for all $I' \in \hat{\mathcal{I}}$. Since $I' \succeq^* I$ for all $I' \in \hat{\mathcal{I}}$, the transitivity of $\succeq^*$ implied by consistency yields $\cup\hat{\mathcal{I}} \succeq^* I$, so $I \cong \cup\hat{\mathcal{I}}$ by (D1).

For (C3), first note that $I \cong I$, so $I \subseteq \pi(I)$. Hence, trivially,

$$\pi(I) \subseteq \pi(I') \implies I \subseteq \pi(I').$$

For the converse, suppose $I \subseteq \pi(I')$. By (D2), then $\pi(I') \succeq^* I$. But (C2) implies $I \cong \pi(I)$, so $I \succeq^* \pi(I)$. Hence by transitivity of $\succeq^*$, $\pi(I') \succeq^* \pi(I)$. Obviously, though, $\pi(I') \succeq^* \pi(I')$, so (C1) implies $\pi(I') \succeq^* \pi(I') \cup \pi(I)$. (D2) implies the reverse direction for $\succeq^*$, so $\pi(I) \cup \pi(I') \cong \pi(I')$ by (D1). Since $I' \cong \pi(I')$ by (C2), transitivity of $\cong$ gives $\pi(I) \cup \pi(I') \cong I'$. Hence the definition of $\pi(I')$ implies

$$\pi(I) \cup \pi(I') \subseteq \pi(I')$$

which implies $\pi(I) \subseteq \pi(I')$.

To show (C4), note that (C3) implies

$$\pi(I) \subseteq \bigcap\{\pi' \in \Pi \mid I \subseteq \pi'\}.$$

Also, note that $I \cong I$, so $I \subseteq \pi(I)$. Hence

$$\pi(I) \in \{\pi' \in \Pi \mid I \subseteq \pi'\},$$

giving the reverse inclusion.

To show (C5), note that $s \cong s$ implies $s \subseteq \pi(s)$. By (D3), $\pi(s) \in \mathcal{I}$. Since S is the set of maximal elements of $\mathcal{I}$, this implies $s = \pi(s)$, so $S \subseteq \Pi$.

For (C6), (C2) implies $I \cong \pi(I)$. Hence

$$\pi(\pi(I)) = \bigcup\{I' \in \mathcal{I} \mid I' \cong \pi(I)\} = \bigcup\{I' \in \mathcal{I} \mid I' \cong I\} = \pi(I).$$

Finally, for (C7), from the definition of $\succeq^*$, $I \cong I \cup I'$ implies $I \succeq^* I'$. To show the converse, suppose $I \succeq^* I'$. Then either $I \cong I \cup I'$ (in which case we are done) or $I \cong I'$. But in the latter case, $I \cong I$ and (C2) imply $I \cong I \cup I'$, so again we are done. ▌

B Proof of Theorem 1

Suppose S^* preserves $\cong$. Then

$$\begin{aligned} I \succeq^* I' &\iff I \cong I \cup I' \text{ or } I \cong I' \\ &\iff S^*(I) = S^*(I \cup I') = S^*(I) \cap S^*(I') \text{ or } S^*(I) = S^*(I') \\ &\iff S^*(I) \subseteq S^*(I') \end{aligned}$$

To show the converse, suppose

$$I \succeq^* I' \iff S^*(I) \subseteq S^*(I').$$

Then (D1) implies

$$\begin{aligned} I \cong I' &\iff I \succeq^* I' \succeq^* I \\ &\iff S^*(I) = S^*(I') \end{aligned}$$

so S^* preserves $\cong$. ▌

C Proof of Theorem 2

First, I show that consistency is necessary. So suppose there exists an S^* preserving $\cong$. By Theorem 1, we know that

$$I \succeq^* I' \iff S^*(I) \subseteq S^*(I').$$

So suppose $I \succeq^* I' \succeq^* I''$. Then we must have $S^*(I) \subseteq S^*(I') \subseteq S^*(I'')$, implying $S^*(I) \subseteq S^*(I'')$. Hence $I \succeq^* I''$, so $\succeq^*$ must be transitive. Also,

$$\begin{aligned} I \succeq^* I' &\iff S^*(I) \subseteq S^*(I') = \textstyle\bigcap_{\varphi \in I'} S^*(\{\varphi\}) \\ &\iff S^*(I) \subseteq S^*(\{\varphi\}), \quad \forall \varphi \in I' \\ &\iff I \succeq^* \{\varphi\}, \quad \forall \varphi \in I' \end{aligned}$$

Hence $\succeq^*$ is consistent.

The proof of sufficiency is by construction. I show that if $\{\succ_I\}$ is consistent, then Π (defined in Appendix A) preserves $\cong$. This is implied by Theorem 6, but I give a direct proof here. By (C3) of Lemma 1, consistency implies

$$\{\pi \in \Pi \mid I \subseteq \pi\} = \{\pi \in \Pi \mid \pi(I) \subseteq \pi\}.$$

Also, by definition of $\pi(I)$, if $I \cong I'$, then $\pi(I) = \pi(I')$. Hence $I \cong I'$ implies

$$\{\pi \in \Pi \mid \pi(I) \subseteq \pi\} = \{\pi \in \Pi \mid \pi(I') \subseteq \pi\}.$$

Conversely, if we have I and I' such that this equality holds, then (C4) of Lemma 1 implies that $\pi(I) = \pi(I')$. By (C2) of Lemma 1, then, $I \cong \pi(I) = \pi(I') \cong I'$, so $I \cong I'$. That is, if we set $S^* = \Pi$, we have $I \cong I'$ if and only if $S^*(I) = S^*(I')$. By (C5), $S \subseteq \Pi$ so Π preserves $\cong$. ▌

D Characterization of the Subjective State Set

Definition 13 $\hat{\Pi} \subseteq \Pi$ *is* unionable *if for all $I \in \mathcal{I}$ such that $I \subseteq \cup\hat{\Pi}$, we have $\pi(I) \subseteq \cup\hat{\Pi}$.*

Let

$$\mathcal{U} = \{\hat{\pi} \subseteq \Phi \mid \hat{\pi} = \cup\hat{\Pi}, \text{ for some unionable } \hat{\Pi} \subseteq \Pi\}.$$

Also, let

$$N = \{\varphi \in \Phi \mid S(\{\varphi\}) = \emptyset\}.$$

That is, N is the set of contradictions.

Definition 14 *A collection of pieces of information $s^* \subseteq \Phi$ is a* copy *of a collection $s' \subseteq \Phi$ if $s^* \setminus N = s'$.*

Note that the definition is written so that a state is a copy of itself. Let $\mathcal{U}^c$ denote the set of copies of elements of $\mathcal{U}$.

Definition 15 $S^* \subseteq 2^\Phi$ spans Π *if for every* $\pi \in \Pi$, *there exists* $\hat{S}^* \subseteq S^*$ *such that*

$$\left[\cap \hat{S}^*\right] \setminus N = \pi.$$

In other words, S^* spans Π if each element of Π is copied by some intersection of elements of S^*.

Theorem 6 *If* $\cong$ *is consistent, then* $S^* \supseteq S$ *preserves* $\cong$ *if and only if* $S^* \subseteq \mathcal{U}^c$ *and spans* Π.

Proof. Suppose $\cong$ is consistent and is preserved by S^*. First, I will show $S^* \subseteq \mathcal{U}^c$. By (C2) for any information set I, $I \cong \pi(I)$, so we must have $S^*(I) = S^*(\pi(I))$. Hence if $I \subseteq s^* \in S^*$, then $\pi(I) \subseteq s^*$.

Clearly, for any $s^* \in S^*$,

$$s^* = \left[\bigcup_{\varphi \in s^* \setminus N} \{\varphi\}\right] \cup \left[\bigcup_{\varphi \in s^* \cap N} \{\varphi\}\right].$$

But each $\varphi \in s^* \setminus N$ is an information set. Hence we must have $\pi(\{\varphi\}) \subseteq s^*$ for each such φ. Since $\varphi \in \pi(\{\varphi\})$ and $\pi(\{\varphi\}) \cap N = \emptyset$ for every information set $\{\varphi\}$,

$$s^* = \left[\bigcup_{\varphi \in s^* \setminus N} \pi(\{\varphi\})\right] \cup \left[\bigcup_{\varphi \in s^* \cap N} \{\varphi\}\right]$$

so that every $s^* \in S^*$ can be written as a copy of a union of sets in Π. Suppose, though, that

$$\hat{\Pi} = \{\pi \in \Pi \mid \pi = \pi(\{\varphi\}), \text{ for some } \varphi \in s^* \setminus N\}$$

is not unionable. Then by definition, there is an $I \subseteq s^*$ such that $\pi(I) \not\subseteq s^*$, contradicting S^* preserving $\cong$. Hence $S^* \subseteq \mathcal{U}^c$.

Next, I show that S^* spans Π. So suppose not. Let π^* be any element of Π such that there is no $\hat{S}^* \subseteq S^*$ with $[\cap \hat{S}^*] \setminus N = \pi^*$. Let $I = \cap S^*(\pi^*) \setminus N$. Clearly, $I \neq \pi^*$. By definition, $s^* \in S^*(\pi^*)$ if and only if $\pi^* \subseteq s^*$. Hence $\pi^* \subseteq \cap S^*(\pi)$. Since $\pi^* \cap N = \emptyset$, $\pi^* \neq I$ implies $\pi^* \subset I$ (where this denotes strict inclusion).

By construction of I, $I \subseteq s^*$ for all $s^* \in S^*(\pi^*)$, so $S^*(\pi^*) \subseteq S^*(I)$. Also, $\pi^* \subset I \subseteq s^*$ for all $s^* \in S^*(I)$, so $S^*(I) \subseteq S^*(\pi^*)$. Hence $S^*(I) = S^*(\pi^*)$. This implies $S(I) = S(\pi^*) \neq \emptyset$, so $I \in \mathcal{I}$. Since S^* preserves $\cong$, this implies $\pi^* \cong I$. Hence the definition of $\pi(\cdot)$ implies $\pi(\pi^*) = \pi(I)$ or, by (C6), $\pi^* = \pi(I)$. But $I \cong I$ implies $I \subseteq \pi(I) = \pi^*$, a contradiction. Hence S^* must span Π.

To show the converse, suppose $S^* \subseteq \mathcal{U}^c$ and that S^* spans Π. I show that this implies that S^* preserves $\cong$. First, fix any I and I' with $I \cong I'$. Fix any $s^* \in S^*(I)$, so $I \subseteq s^*$. Because $I \cap N = \emptyset$, we must have $I \subseteq s^* \setminus N$. Because $s^* \setminus N \in \mathcal{U}$, $I \subseteq s^* \setminus N$ implies $\pi(I) \subseteq s^* \setminus N$. But $I' \cong I$ implies $I' \subseteq \pi(I)$, so

$$I' \subseteq \pi(I) \subseteq s^* \setminus N \subseteq s^*.$$

Hence $S^*(I) \subseteq S^*(I')$. The reverse inclusion follows from reversing the roles of I and I' so $S^*(I) = S^*(I')$. Hence $I \cong I'$ implies $S^*(I) = S^*(I')$.

To show that the reverse implication holds as well, suppose we have I and I' with $S^*(I) = S^*(I')$. Obviously, then,

$$\{\pi \in \Pi \mid \exists \hat{S}^* \subseteq S^*(I) \text{ with } \pi = [\cap \hat{S}^*] \setminus N\} = \{\pi \in \Pi \mid \exists \hat{S}^* \subseteq S^*(I') \text{ with } \pi = [\cap \hat{S}^*] \setminus N\}.$$

Clearly, if $\hat{S}^* \subseteq S^*(I)$, then $I \subseteq \cap \hat{S}^*$. Hence any π in the set on the left must have $I \subseteq \pi$. Furthermore, if we have a π which contains I, the only way to find a set $\hat{S}^*$ satisfying $[\cap S^*] \setminus N = \pi$ is by taking $\hat{S}^* \subseteq S^*(I)$. Hence the set on the left is precisely the set of π containing I which S^* spans. Since S^* spans Π, this must be all the π's containing I. In short, this is equivalent to

$$\{\pi \in \Pi \mid I \subseteq \pi\} = \{\pi \in \Pi \mid I' \subseteq \pi\}$$

so

$$\bigcap \{\pi \in \Pi \mid I \subseteq \pi\} = \bigcap \{\pi \in \Pi \mid I' \subseteq \pi\}.$$

By (C4), this is equivalent to $\pi(I) = \pi(I')$. Hence $I \cong I'$. Hence S^* preserves $\cong$. ▌

E Proof of Theorem 3

The necessity of representability is obvious. Theorem 2 shows that consistency is necessary. The sufficiency proof is by construction. First, suppose that $\{\succ_I\}$ satisfies representability and consistency and Φ is broad. For each $s \in S$, fix a copy s' of s with $s' \neq s$. (The broadness of Φ implies that this is possible.) Let the collection of these sets be denoted S'. Let the extended state set be $S^* = \Pi \cup S'$. Since S is finite, S' is finite. By Assumption 2, Π is finite, so S^* is finite as well. By (C5) of Lemma 1, $S \subseteq \Pi$, so this specification guarantees that $S \subseteq S^*$, as required. By Theorem 6, this specification of S^* satisfies the requirement that S^* preserve $\cong$.

To continue the construction of the Extended EU representation, let $u(x) = x$ for all $x \in \mathbf{R}$. Let μ be uniform on S^*. The final object to construct is the extension function h. To do this, first fix a collection of utility functions $u_\pi : F \to \mathbf{R}$, one for each $\pi \in \Pi$, where u_π represents $\succ_\pi$. By representability, this is possible for all nonnull π. If π is null, then $f \sim_\pi g$ for all $f, g \in F$, so, obviously, again it is possible to find a u_π function.

As required, let $h(f)(s) = f(s)$ for all $s \in S$. For $s' \in S'$, let

$$h(f)(s') = u_{s' \setminus N}(f) - f(s' \setminus N).$$

Recall that $s' \setminus N \in S$ for all $s' \in S'$.

For $\pi \in \Pi \setminus S$, we define $h(f)(\pi)$ by induction. To set up this induction, let $\Pi_0 = S$ and for $k \geq 1$, let

$$\Pi_k = \left\{ \pi \in \Pi \setminus \bigcup_{j=0}^{k-1} \Pi_j \mid \pi \subset \pi' \text{ implies } \pi' \in \bigcup_{j=0}^{k-1} \Pi_j \right\}$$

where $\subset$ refers to strict inclusion. Because elements of Π are information sets, each must be contained in some $s \in S$. Furthermore, since S is the set of maximal elements of $\mathcal{I}$, no $s \in S$ can be contained in any element of Π except itself. By the finiteness of Π, there must be some π's which are strictly contained only by π''s in S. It is easy to generalize this to show that there is some smallest finite K such that $\cup_{k=0}^{K} \Pi_k = \Pi$ and that every $\pi \in \Pi$ is contained in Π_k for a unique k. Let $k(\pi)$ denote that k such that $\pi \in \Pi_k$.

We have already defined $h(f)(\pi)$ for π with $k(\pi) = 0$. To complete the specification of h, let

$$h(f)(\pi) = u_\pi(f) - \sum_{\pi' \in \Pi \mid \pi \subset \pi'} h(f)(\pi') - \sum_{s' \in S' \mid \pi \subset s'} h(f)(s').$$

Since $\pi \subset \pi'$ implies $k(\pi) > k(\pi')$, this is a legitimate definition by induction.

We now show that this specification of u, μ, S^*, and h guarantees that we represent the preferences. First, notice that if $I \subseteq s^* \in S^*$, then $\pi(I) \subseteq s^*$. To see this, notice that it is implied by (C3) if $s^* \in \Pi$. If $s^* \in S'$, then $I \subseteq s^*$ if s^* is a copy of an element of $S(I)$. But $\pi(I) \cong I$, so $S(I) = S(\pi(I))$. Hence s^* is a copy of an element of $S(\pi(I))$, so $\pi(I) \subseteq s^*$.

Hence for any $I \in \mathcal{I}$ and any $f \in F$, we have

$$\begin{aligned}
\mathrm{E}_\mu[u(h(f)(s^*)) \mid s^* \in S^*(I)] &= \tfrac{1}{\#S^*} \textstyle\sum_{s^* \in S^*(I)} h(f)(s^*) \\
&= \tfrac{1}{\#S^*} [u_{\pi(I)}(f) - \textstyle\sum_{\pi' \in \Pi \mid \pi(I) \subset \pi'} h(f)(\pi') \\
&\qquad - \textstyle\sum_{s' \in S' \mid \pi(I) \subset s'} h(f)(\pi'') \\
&\qquad + \textstyle\sum_{\pi' \in \Pi \mid \pi(I) \subset \pi'} h(f)(\pi') \\
&\qquad + \textstyle\sum_{s' \in S' \mid \pi(I) \subset s'} h(f)(\pi'')] \\
&= \tfrac{1}{\#S^*} u_{\pi(I)}(f).
\end{aligned}$$

This function represents $\succ_I$ since $u_{\pi(I)}$ represents $\succ_{\pi(I)}$ and $\pi(I) \cong I$ implies $\succ_{\pi(I)} = \succ_I$. Since $\mu(S(I)) > 0$ for all I, it is certainly positive for nonnull I.

Now replace the broadness of Φ with the assumption that $\{\succ_I\}$ satisfies weak state independence. Let u be the function WSI requires. Let $S^* = \Pi$. By Assumption 2, this is finite. Obviously, Theorem 6 implies that this preserves $\cong$. Let

$$\bar{S}^* = \{s^* \in S^* \mid s^* \not\cong s, \text{ for any null } s \in S\}.$$

Let $\mu(\pi) = 1/\#\bar{S}^*$ for all $\pi \in \bar{S}^*$ and $\mu(\pi) = 0$ for $\pi \in S^* \setminus \bar{S}^*$. For each $\pi \in \bar{S}^* \setminus S$, let $u_\pi : F \to \mathbf{R}$ be a utility function representing $\succ_\pi$. Just as in the first sufficiency proof, this is obviously possible for null π and is guaranteed possible for nonnull π by representability of **P**.

To define h, set $h(f)(s) = f(s)$ for $s \in S$. For $\pi \in S^* \setminus \bar{S}^*$, $h(f)(\pi)$ can be chosen arbitrarily. Finally, for $\pi \in \bar{S}^*$, I follow a procedure analogous to that used in the first sufficiency proof. In particular, let

$$u(h(f)(\pi)) = u_\pi(f) - \sum_{\pi' \in \bar{S}^* | \pi \subset \pi'} u(h(f)(\pi')).$$

Since u is onto and since $\pi \subset \pi'$ implies $k(\pi) > k(\pi')$ (where k is defined as above), this is a legitimate definition by induction.

We now show that this specification of u, μ, S^*, and h guarantees that we represent the preferences. As before, $I \subseteq \pi \in S^*$ implies $\pi(I) \subseteq \pi$ by (C3). Hence for any $f \in F$ and any $I \in \mathcal{I}$ with $I \not\cong s$ for any s, we have

$$\begin{aligned} \mathrm{E}_\mu[u(h(f)(s)) \mid s^* \in S^*(I)] &= \tfrac{1}{\#\bar{S}^*} \textstyle\sum_{s^* \in \bar{S}^*} u(h(f)(s^*)) \\ &= \tfrac{1}{\#\bar{S}^*} [u_{\pi(I)}(f) - \textstyle\sum_{\pi' \in \bar{S}^* | \pi(I) \subset \pi'} u(h(f)(\pi')) \\ &\qquad + \textstyle\sum_{\pi' \in \bar{S}^* | \pi(I) \subset \pi'} u(h(f)(\pi'))] \\ &= \tfrac{1}{\#\bar{S}^*} u_{\pi(I)}(f). \end{aligned}$$

This function represents $\succ_I$ since $u_{\pi(I)}$ represents $\succ_{\pi(I)}$ and $\pi(I) \cong I$ implies $\succ_{\pi(I)} \; = \; \succ_I$. For any nonnull I with $I \cong s$ for some s, $S^*(I) = \{s\}$. Hence for any $f \in F$,

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

By weak state independence, this represents $\succ_I$. Hence $\succ_I$ is represented for every nonnull I. Clearly, $\mu(S(I)) > 0$ for all nonnull I. ▌

F Proof of Theorem 4

In the remainder of the Appendix, when referring to a singleton information set $\{\varphi\}$, I often omit the braces for notational ease.

First consider any singleton information set φ. Because $\{\succ_I\}$ respects simple implication, (C7) implies $\varphi \cong \{\varphi, \psi\}$ for any ψ with $S(\varphi) \subseteq S(\psi)$. Hence

$$\{\psi \in \Phi \mid S(\varphi) \subseteq S(\psi)\} \subseteq \pi(\varphi).$$

But $S(\varphi) \subseteq S(\psi)$ if and only if ψ is true in every state in $S(\varphi)$ — that is, if and only if $\psi \in \cap S(\varphi)$. Hence

$$\{\psi \in \Phi \mid S(\varphi) \subseteq S(\psi)\} = \cap S(\varphi) \subseteq \pi(\varphi).$$

To see that the inclusion cannot be strict, consider any $\psi \notin \cap S(\varphi)$. Then there is a state $s \in S(\varphi)$ with $\psi \notin s$. But then for any information set I containing ψ, $s \notin S(I)$. Hence for any such information set, $S(I) \neq S(\varphi)$, so $I \ncong \varphi$, implying $\psi \notin \pi(\varphi)$. Hence for any singleton information set φ, we must have $\pi(\varphi) = \cap S(\varphi)$.

It is easy to see that if $\cong$ is consistent, then for any information set I,

$$I \cong \bigcup_{\varphi \in I} \pi(\varphi).$$

This holds because if $\cong$ is consistent, there is some S^* which preserves it. For this S^*,

$$S^*(I) = \bigcap_{\varphi \in I} S^*(\varphi) = \bigcap_{\varphi \in I} S^*(\pi(\varphi)) = S^*(\bigcup_{\varphi \in I} \pi(\varphi)),$$

proving the claim. In particular, for any $\pi \in \Pi$,

$$\pi \cong \bigcup_{\varphi \in \pi} \pi(\varphi).$$

Hence by the definition of π and using (C6),

$$\bigcup_{\varphi \in \pi} \pi(\varphi) \subseteq \pi(\pi) = \pi.$$

Obviously, though,

$$\pi = \bigcup_{\varphi \in \pi} \{\varphi\} \subseteq \bigcup_{\varphi \in \pi} \pi(\varphi).$$

Hence for each $\pi \in \Pi$,

$$\pi = \bigcup_{\varphi \in \pi} \pi(\varphi) = \bigcup_{\varphi \in \pi} \cap S(\varphi) \in \tau.$$

Hence $\Pi \subseteq \tau$. Because any union of elements of τ is also an element of τ, this implies $\mathcal{U} \subseteq \tau$. Hence by Theorem 6 and Assumption 3, $S^* \subseteq \tau$.

For the converse, suppose $S^* \subseteq \tau$. Suppose $S(\varphi) \subseteq S(\psi)$. Consider any $s^* \in S^*(\varphi)$. Since $s^* \in \tau$, there must be a collection of subsets of S, say $\mathcal{B}$, such that

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

Since $\varphi \in s^*$, it must be true that $\varphi \in \cap B$ for some $B \in \mathcal{B}$. Hence there is a $B \in \mathcal{B}$ with $B \subseteq S(\varphi)$. But then $S(\varphi) \subseteq S(\psi)$ implies $B \subseteq S(\psi)$, so $\psi \in \cap B$ and $\psi \in s^*$. Hence $S^*(\varphi) \subseteq S^*(\psi)$. Since S^* preserves $\cong$, this implies that if $S(\varphi) \subseteq S(\psi)$, we have $\{\varphi\} \succeq^* \{\psi\}$, so $\{\succ_I\}$ preserves simple implication. ▌

G Proof of Theorem 5

Assume $\{\succ_I\}$ respects logical equivalence so $I \cong I'$ if and only if $S(I) = S(I')$. I first show that this implies that

$$\pi(I) = \{\varphi \in \Phi \mid S(I) \subseteq S(\varphi)\}.$$

To see this, note that $\varphi \in \pi(I)$ iff there is an I' with $\varphi \in I'$ and $I' \cong I$. Hence there is an I' containing φ with

$$S(I) = S(I') = \bigcap_{\psi \in I'} S(\psi) \subseteq S(\varphi).$$

Hence $\varphi \in \pi(I)$ implies $S(I) \subseteq S(\varphi)$. For the converse, suppose $S(I) \subseteq S(\varphi)$. Then

$$S(I \ \cup \ \{\varphi\}) = S(I) \ \cap \ S(\varphi) = S(I),$$

so $I \cong I \ \cup \ \{\varphi\}$, implying $\varphi \in \pi(I)$.

Therefore,

$$\pi(I) = \{\varphi \in \Phi \mid S(I) \subseteq S(\varphi)\} = \{\varphi \in \Phi \mid \varphi \in \cap S(I)\} = \cap S(I).$$

By richness of Φ, for every nonempty $B \subseteq S$, there is a φ_B with $S(\varphi_B) = B$. Hence for every $B \subseteq S$, there is an information set I such that $\pi(I) = B$. So $\Pi = S^*_\cap$.

To complete the characterization of S^*, I need to characterize the unionable subsets of Π. So suppose $\hat{\Pi} \subseteq \Pi$ is unionable. We know that $\hat{\Pi}$ is always unionable if $\cup\hat{\Pi} = \cup\Pi$, so suppose $\cup\hat{\Pi} \subset \cup\Pi$. We know that $\hat{\Pi}$ is unionable if $\cup\hat{\Pi} \in \Pi$, so assume this is not the case. Hence, in particular, $\hat{\Pi}$ contains at least two elements. From the result above,

$$\cup\hat{\Pi} = \bigcup_{\pi \in \hat{\Pi}} \cap S(\pi).$$

Clearly, if there exists $\pi_1, \pi_2 \in \hat{\Pi}$ with $S(\pi_1) \subseteq S(\pi_2)$, then $\cap S(\pi_1) \supseteq \cap S(\pi_2)$. Hence $\cup\hat{\Pi}$ is unaffected by eliminating any such π_2, so, without loss of generality, I assume there is no such π_1 and π_2.

First, suppose we have $\pi_1, \pi_2 \in \hat{\Pi}$ with $S(\pi_1) \ \cap \ S(\pi_2) \neq \emptyset$ but $S(\pi_1) \not\subseteq S(\pi_2)$ and vice versa. By richness of Φ, there exists φ_1 and φ_2 such that $S(\varphi_i) = S(\pi_i)$ for $i = 1, 2$. Hence the fact that $\{\succ_I\}$ preserves logical equivalence implies $\{\varphi_i\} \cong \pi_i$, so $\varphi_i \in \pi_i$ for $i = 1, 2$. Therefore, $\{\varphi_1, \varphi_2\} \subseteq \cup\hat{\Pi}$. Also, $S(\{\varphi_1, \varphi_2\}) = S(\varphi_1) \cap S(\varphi_2) = S(\pi_1) \cap S(\pi_2)$ which is nonempty by assumption. Hence $\{\varphi_1, \varphi_2\}$ is an information set contained in $\cup\hat{\Pi}$. By the definition of unionability, then,

$$\pi(\{\varphi_1, \varphi_2\}) \subseteq \cup\hat{\Pi} = \bigcup_{\pi \in \hat{\Pi}} \cap S(\pi).$$

But the fact that $\{\succ_I\}$ respects logical equivalence implies

$$\pi(\{\varphi_1, \varphi_2\}) = \cap S(\{\varphi_1, \varphi_2\}) = \cap\left[S(\pi_1)\ \cap S(\pi_2)\right].$$

Clearly, $S(\pi_1)\ \cap\ S(\pi_2) \subset S(\pi_i)$ for each i, so $\cap[S(\pi_1)\ \cap\ S(\pi_2)] \supseteq \cap S(\pi_i)$. Hence $\cup\hat{\Pi}$ remains unchanged and its unionability unaffected if replace π_1 and π_2 with $\pi_1\ \cup\ \pi_2$ (since $S(\pi_1\ \cup\ \pi_2) = S(\pi_1)\ \cap\ S(\pi_2)$). Hence we can assume that for every $\pi_1,\ \pi_2 \in \hat{\Pi}$, $S(\pi_1)\ \cap\ S(\pi_2) = \emptyset$.

So suppose that we have $\pi_1, \pi_2 \in \hat{\Pi}$ with $S(\pi_1)\ \cap\ S(\pi_2) = \emptyset$. Suppose $\#S(\pi_2) \geq 2$. Fix any $s \in S(\pi_2)$. By richness of Φ, there exists φ with $S(\varphi) = S(\pi_1)\ \cup\ \{s\}$. Hence $\varphi \in \cap S(\pi_1)$, so $\varphi \in \cup\hat{\Pi}$. Note that

$$S(\pi_2\ \cup\ \{\varphi\}) = S(\pi_2)\ \cap\ S(\varphi) = S(\pi_2)\ \cap\ \left[S(\pi_1) \setminus \{s\}\right] = \{s\}.$$

Hence $\pi_2\ \cup\ \{\varphi\}$ is an information set and is contained in $\cup\hat{\Pi}$. By unionability, then, $\pi(\pi_2\ \cup\ \{\varphi\}) \subseteq \hat{\Pi}$. However, the fact that $\{\succ_I\}$ respects logical equivalence implies that $\pi_2\ \cup\ \{\varphi\} \cong s$. Hence we have

$$\pi(\pi_2\ \cup\ \{\varphi\}) = s \subseteq \bigcup_{\pi\in\hat{\Pi}} \cap S(\pi).$$

Clearly, $s \in S(\pi_2)$ implies $s \supseteq \cap S(\pi_2)$. Hence $\cup\hat{\Pi}$ remains unchanged and its unionability unaffected if we assume that for every $\pi \in \hat{\Pi}$, $\#S(\pi) = 1$.

Therefore, if $\hat{\Pi}$ is unionable, then, unless $\cup\hat{\Pi} \in \Pi$ or $\cup\hat{\Pi} = \cup\Pi$, we must have

$$\cup\hat{\Pi} = \cup B$$

for some $B \subseteq S$. Since every $\pi \in \Pi$ is an information set, $\cup\Pi \subseteq \cup S$. Since $S \subseteq \Pi$ by (C5), we must have $\cup\Pi = \cup S$. Hence $\cup\hat{\Pi} \neq \cup\Pi$ implies $B \neq S$. Furthermore, $S \subseteq \Pi$ and $\cup\hat{\Pi} \notin \Pi$ implies $\#B \geq 2$.

Fix any $s \in S \setminus B$ and any $s_1, s_2 \in B$. By richness of Φ, there exist φ_1 and φ_2 such that $S(\varphi_i) = \{s, s_i\}$, for $i = 1, 2$. Clearly, $\varphi_i \in \cup\hat{\Pi}$ for $i = 1, 2$, so $\{\varphi_1, \varphi_2\}$ is an information set contained in $\cup\hat{\Pi}$. By unionability, then,

$$\pi(\{\varphi_1, \varphi_2\}) \subseteq \cup\hat{\Pi}.$$

But $\pi(\{\varphi_1, \varphi_2\}) = s$. Since s was arbitrary, this implies $\cup S \subseteq \cup\hat{\Pi}$. As noted above, $\cup\Pi = \cup S$, so this implies $\cup\Pi = \cup\hat{\Pi}$.

Hence the only unionable subsets of Π are subsets $\hat{\Pi}$ such that $\cup\hat{\Pi} \in \Pi$ or $\cup\hat{\Pi} = \cup\Pi = \cup S$. Because we have assumed that there are no contradictions in Φ, $\cup S = \Phi$. Hence $\mathcal{U} = S^*_\cap\ \cup\ [\cup S]$, so by Theorem 6, $S^* \subseteq S^*_\cap\ \cup \{\Phi\}$.

For the converse, suppose $S^* \subseteq S^*_\cap \cup \{\Phi\}$. Suppose $S(I) = S(I')$. Fix any $s^* \in S^*(I)$. Clearly, if $s^* = \Phi$, then $I' \subseteq s^*$ as well. So suppose $s^* \neq \Phi$. Then there is $B \subseteq S$ such that $s^* = \cap B$. Since $I \subseteq s^*$, we must have $B \subseteq S(I)$. But then $S(I) = S(I')$ implies $B \subseteq S(I')$, so $I' \subseteq s^*$ also. Therefore, $S^*(I) \subseteq S^*(I')$. The symmetric argument establishes the converse, so $S^*(I) = S^*(I')$. Since S^* preserves $\cong$, we see that $\{\succ_I\}$ respects logical equivalence. ▌

References

[1] Chew, S. H., and E. Karni, "Choquet Expected Utility with Finite State Space: Commutativity and Act–Independence," *Journal of Economic Theory*, 62, April 1994, 469–479.

[2] Dekel, E., B. Lipman, and A. Rustichini, "Recent Developments in Modeling Unforeseen Contingencies," working paper, 1997.

[3] Fagin, R., J. Halpern, Y. Moses, and M. Vardi, *Reasoning about Knowledge*, Cambridge: MIT Press, 1995.

[4] Fishburn, P., "Intransitive Indifference in Preference Theory: A Survey," *Operations Research*, 10, 1973, 207–228.

[5] Ghirardato, P., and M. Le Breton, "Choquet Rationalizability," Cal Tech working paper, March 1997.

[6] Gilboa, I., and D. Schmeidler, "Additive Representations of Non–Additive Measures and the Choquet Integral," *Annals of Operations Research*, 52, 1994, 43–65.

[7] Gul, F., "Savage's Theorem with a Finite Number of States," *Journal of Economic Theory*, 57, June 1992, 99–110.

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

[9] Kreps, D., "A Representation Theorem for 'Preference for Flexibility,'" *Econometrica*, 47, May 1979, 565–577.

[10] Kreps, D., "Static Choice in the Presence of Unforeseen Contingencies," in *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, P. Dasgupta, D. Gale, O. Hart, and E. Maskin (eds.), Cambridge: MIT Press, 1992, 259–281.

[11] Lipman, B., "How to Decide How to Decide How to . . . : Modeling Limited Rationality," *Econometrica*, 59, July 1991, 1105–1125.

[12] Lipman, B., "Logics for Nonomniscient Agents: An Axiomatic Approach," forthcoming in M. Bacharach, L. A. Gerard-Varet, P. Mongin, H. Shin, eds., *Epistemic Logic and the Theory of Games and Decisions*, Dordrecht, Netherlands: Kluwer, 1994a.

[13] Lipman, B., "An Axiomatic Approach to the Logical Omniscience Problem," in R. Fagin, ed., *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Fifth Conference*, San Francisco, CA: Morgan Kaufmann Publishers, 1994b.

[14] Lipman, B., "Information Processing and Bounded Rationality: A Survey," *Canadian Journal of Economics*, 28, February 1995, 42–67.

[15] Luce, R. D., and D. Krantz, "Conditional Expected Utility," *Econometrica*, 39, 1971, 565–577.

[16] Morris, S., "The Logic of Belief and Belief Change: A Decision Theoretic Approach," *Journal of Economic Theory*, 69, April 1996, 1–23.

[17] Mukerji, S., and H. S. Shin, "Equilibrium Departures from Common Knowledge in Games with Non–Additive Expected Utility," working paper, 1997.

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

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

[20] Rubinstein, A., "Comments on the Interpretation of Game Theory," *Econometrica*, 59, July 1991, 909–924.

[21] Rubinstein, A., *Modeling Bounded Rationality,* Cambridge: MIT Press, 1998.

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

[23] Savage, L. J., "Difficulties in the Theory of Personal Probability," *Philosophy of Science*, 34, 1967, 305–310.

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

[25] Shin, H. S., "Logical Structure of Common Knowledge," *Journal of Economic Theory*, 60, 1993, 1–13.

[26] Skiadis, C., "Subjective Probability under Additive Aggregation of Conditional Preferences," working paper, Department of Finance, Northwestern University, 1996.