Econometrica, Vol. 66, No. 1 (January, 1998), 159–173

STANDARD STATE-SPACE MODELS PRECLUDE UNAWARENESS

BY EDDIE DEKEL, BARTON L. LIPMAN, AND ALDO RUSTICHINI[1]

1. INTRODUCTION

ONE ASPECT OF BOUNDED RATIONALITY with important economic implications is unawareness of various possibilities. For example, unforeseen contingencies could prevent agents from writing the kinds of contracts our models predict. In this paper we explore the extent to which commonly used models need to be modified in order to capture unawareness.

We start by examining information structures that generalize the standard partitional model. These structures, called *possibility correspondences*, are often motivated by bounded rationality in general and unawareness in particular (see, for example, Brandenburger, Dekel, and Geanakoplos (1992), Geanakoplos (1989), Morris (1996), Morris and Shin (1992), Samet (1990), and Shin (1993)). Formally, a possibility correspondence is a function P from the set of states, Ω, to subsets of Ω, where $P(\omega)$ is interpreted as the set of states the agent considers possible when the true state is ω. Thus the agent is said to know an event E at state ω if $P(\omega) \subseteq E$. A possibility correspondence is *partitional* if its image partitions Ω and is *nonpartitional* otherwise.

We illustrate the intuition behind the connection between unawareness and non-partitional possibility correspondences in Example 1 below. A connection can also be seen via one of the necessary conditions for a possibility correspondence to be partitional; namely, that it satisfy a property called *negative introspection* (also called *knowing that you don't know*). This axiom says that in any state where the agent doesn't know some event, she does know that she doesn't know it. For example, suppose a person contracting to have a home built fails to foresee the possibility that city regulations preclude locating his driveway where he wants it.[2] It seems quite natural to say that this person does not know that city regulations will preclude the planned driveway location. But if this possibility is truly unforeseen, it seems absurd to say that the person knows that he doesn't know about these regulations. Thus the possibility correspondences that are said to model unawareness are those that violate negative introspection and hence are nonpartitional.

To explore whether possibility correspondences can model unawareness, we consider an unawareness operator which specifies, for each state of the world, the events in the state space of which the agent is unaware. The only conditions we put on this operator

[1] The authors thank Ariel Rubinstein, Dov Samet, Tim van Zandt, Guy Laroque, and three anonymous referees for comments and Tel-Aviv University for its hospitality during part of the work on this paper. Dekel and Rustichini thank the NSF (Grants SBR 9409302 and SBR 9514908 respectively) and Lipman thanks SSHRCC for financial support for this research. Dekel and Lipman particularly thank Phil Reny for a series of discussions which led to this project. Rustichini thanks Salvatore Modica for long, illuminating discussions on the topic of the paper. This paper was formerly titled "Possibility Correspondences Preclude Unawareness."

[2] Hart (1995) gives this as an example of a contingency he himself did not foresee in the planned construction of his home.

are three axioms, all based on the idea that unawareness of a possibility corresponds to a complete lack of positive knowledge regarding it. Our first axiom is that this lack of positive knowledge *at least* includes the first two levels—that is, we require that an agent who is unaware of an event at least does not know that the event is true and does not know that he does not know. The other two axioms extend this lack of positive knowledge to knowledge about unawareness itself. The second axiom is that the agent can never know that he is unaware of any given specific event. Finally, we assume that if the agent is unaware of an event E, then he is unaware of the possibility of being unaware of E.

Our first main result shows that in possibility correspondence models, any unawareness operator satisfying the three axioms must be trivial.[3] On the other hand, as we show by example, outside the class of possibility correspondence models, one can have a nontrivial unawareness operator satisfying our axioms. To evaluate this example, we turn in Section 3 to the underpinnings of the state-space model. In doing so, we discover an additional property that any appealing concept of unawareness should satisfy. The property, which we call *weak necessitation*, says that awareness of some proposition φ should imply knowledge that some tautology concerning φ is true. For example, if you are aware of φ, then you should know that φ implies φ. Our second main result is that the three axioms combined with the new property preclude unawareness in any "standard" state-space model.

Stated rather vaguely, the intuition behind our results is very simple: if an agent is unaware of some possibility, he must not fully understand the state space. The way we usually work with state-space models requires the agent to have more understanding of the state space than unawareness allows. By exploring the underpinnings of the state-space model, we are also able to state precisely how much understanding is "too much" to allow unawareness. We identify two key assumptions on which "standard" state-space models are based. The first assumption is that all states in the model are real possibilities, as opposed to objects present only to describe the agent's perception of possibilities. The second assumption is that all relevant properties of any given fact about the world can be completely summarized by a subset of the state space, in the sense that whether the agent knows a given fact at a given state is entirely determined by the set of states in which this fact is true.

Can the "standard" state-space model be modified in a way which allows nontrivial unawareness and satisfies all the properties we have proposed? While our main message is a negative one concerning the use of "standard" state spaces to model unawareness, we address this question briefly in the conclusion. We illustrate a resolution of the conflict between knowledge of the model and unawareness with an example, loosely based on Modica and Rustichini (1993), of how we can drop the "real-states" assumption and get a nonstandard state-space model which might be able to capture a more interesting notion of unawareness. The details of the example are contained in the Appendix. We emphasize that our purpose in giving this example is to delineate the boundary of the negative result, not to "prove" that this approach does provide a useful model of unawareness.

The rest of the introduction develops an example, based on one in Geanakoplos (1989), and uses it to present the model, motivate our axioms, explain our results, and discuss related literature. First, we show why many have suggested that possibility correspondences might model unawareness.

[3] The restriction to nontrivial cases is needed because the axioms always hold in trivial cases such as when the agent is aware of everything.

EXAMPLE 1: *Part A*.

While Watson never reported it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not howl, Holmes replied "that is the curious incident to which I refer." Holmes knew that this meant that no one, neither man nor dog, had intruded on the premises the previous night. For had a man intruded, the dog would have barked. Had a dog intruded, the cat would have howled. Hence the lack of either of these two signals means that there could not have been a human or canine intruder.

In a possibility-correspondence structure that might model this story, let a denote the state in which there is a human intruder, b the state in which there is a dog intruder, and c the state where there is no intruder at all. So Ω, the set of all states, is $\{a, b, c\}$. In state a, the dog barks, causing Watson to realize there is a human intruder. In state b, the cat howls, causing Watson to realize that there is a canine intruder. Finally, in state c, neither happens and Watson, not recognizing the significance of this, does not update at all—that is, he still considers states a, b, and c possible. So his possibility correspondence is $P(a) = \{a\}$, $P(b) = \{b\}$, and $P(c) = \{a, b, c\}$. Intuitively, in state c, Watson seems unaware of the possibility that the dog would have barked or the cat would have howled and this is why he fails to recognize that the true state must be c.

Further support for the intuition that possibility correspondences might model unawareness comes from the fact that this instance of apparent unawareness corresponds to a violation of negative introspection, precisely the assumption we suggested a model of unawareness would have to drop. To make this precise, we formally define the associated knowledge operator.

Given a possibility correspondence, we define a *knowledge operator* $K: 2^{\Omega} \to 2^{\Omega}$ by $K(E) = \{\omega \in \Omega \mid P(\omega) \subseteq E\}$. Intuitively, $K(E)$ is the set of states in which Watson knows that E must have occurred—i.e., the set of states such that when they occur he knows that the true state must lie in E. In this example, the set of states where Watson knows that $\{a\}$ has occurred (that is, knows there was a human intruder) is simply $\{a\}$. Hence the set of states where he doesn't know that $\{a\}$ has occurred is the complement of this, or $\{b,c\}$. Using $\neg$ to denote complementation, then, $\neg K(E)$ is the set of states where Watson does not know that E has occurred. When does he know that he doesn't know that E has occurred? That is, when does he recognize his lack of knowledge about E? Since $\neg K(E)$ is the set of states where he doesn't know that E occurred, $K(\neg K(E))$ is the set of states where he knows that he does not know.[4]

So when does Watson know that he doesn't know $\{a\}$? Clearly,

$$K \neg K(\{a\}) = K(\{b, c\}) = \{b\}.$$

Hence

$$\neg K \neg K(\{a\}) = \neg \{b\} = \{a, c\}.$$

So the set of states where he doesn't know $\{a\}$ and also doesn't know that he doesn't know $\{a\}$ would be the intersection of $\{b, c\}$ and $\{a, c\}$, i.e., $\{c\}$. In other words, the state at which we suggested Watson seemed to be unaware of the possibility of the dog barking (something which occurs only when the state is a) is one where negative introspection breaks down. At that state, he doesn't know $\{a\}$ but also doesn't know that he doesn't

[4] Because we will frequently consider iterated knowledge of this sort, we will often omit the parentheses for easier reading.

know. Also, one can easily calculate further levels of not knowing following the same procedure to show that at c, Watson has no positive knowledge regarding $\{a\}$ at all. That is,

$$\{c\} = \bigcap_{i=1}^{\infty} (\neg K)^{i}(\{a\}),$$

where $(\neg K)^i$ means that we iterate the $\neg K$ operation i times. Hence in state c, Watson seems unaware of $\{a\}$.

Why then do we claim that possibility correspondences cannot model unawareness? As above, it seems intuitive that an agent who is unaware of a possibility should have *no* positive knowledge of it at all. In particular, he should not know or be aware that he is unaware of this possibility. Our results show that these properties cannot be satisfied by nonpartitional possibility correspondences (or standard state-space models more generally). To illustrate this, we return to Example 1.

EXAMPLE 1: *Part B*.

Suppose one defines unawareness by letting $U(E)$ be the set of states where Watson is unaware of the event E and requiring that

$$U(E) = \bigcap_{i=1}^{\infty} (\neg K)^{i}(E).$$

It is not hard to show that

$$U(\varnothing) = U(\Omega) = U(\{c\}) = U(\{a,b\}) = \varnothing$$

while

$$U(\{a\}) = U(\{b\}) = U(\{b,c\}) = U(\{a,c\}) = \{c\}.$$

Note that $U(\{a\}) = \{c\}$, while $UU(\{a\}) = U(\{c\}) = \varnothing$. In other words, at c, Watson is unaware of a but is aware that he is unaware of $\{a\}$! While it seems quite plausible to say that Watson knows there is *something* he is unaware of, it seems completely unreasonable for him to be aware of precisely which event he is unaware. Our third axiom rules out this possibility and hence this example.

Our criticisms are similar to those of Modica-Rustichini (1994), but differ in several key respects. Like Modica-Rustichini, we argue that nonpartitional possibility correspondence models are incapable of capturing an interesting form of unawareness. They also identify some of the same properties of unawareness as being of interest. On the other hand, their critique of unawareness is based on a specific and not entirely convincing definition. In particular, they say that an agent is unaware of a possibility if he does not know that the possibility is true and does not know that he does not know. As we show below, this is an odd definition in that these two levels of lack of knowledge do not, in general, imply any higher level lack of knowledge. In addition, their analysis uses an assumption of symmetry in a critical way. Specifically, they assume that an agent is unaware of a possibility if and only if he is unaware of its negation. While not without appeal, symmetry is not obviously an essential property of unawareness; moreover, as our results make clear, it is simply not relevant.[5]

By contrast, we give a very simple construction with weaker and more intuitive hypotheses which enables us to get directly to the heart of the matter. As our main results and their very simple proofs clearly indicate, what is involved is very fundamental and intuitive. This is also why we are able to go beyond the class of possibility correspondences to show that standard state-space models preclude unawareness.

To illustrate our differences with Modica-Rustichini, we return once more to the Watson example.

EXAMPLE 1: *Part C.*

The Modica-Rustichini definition is that an agent is unaware of event E if he doesn't know E and doesn't know that he doesn't know E. That is, using $U_{MR}(E)$ to denote their unawareness operator, we have

$$U_{MR}(E) = \neg K(E) \cap \neg K \neg K(E).$$

It is not hard to show that $U_{MR}(E) = U(E)$ for the events $\{a\}$, $\{b\}$, $\{b, c\}$, and $\{a, c\}$.

On the other hand, consider $\{a, b\}$. It is easy to see that

$$\neg K(\{a, b\}) = \neg\{a, b\} = \{c\}$$

so

$$\neg K \neg K(\{a, b\}) = \neg K(\{c\}) = \Omega.$$

Therefore $U_{MR}(\{a, b\}) = \{c\}$. Hence Modica-Rustichini's symmetry requirement says that since Watson is unaware of $\{a, b\}$ at state c, he must also be unaware of $\{c\}$. However,

$$\neg K(\{c\}) = \Omega$$

so

$$K \neg K(\{c\}) = K(\Omega) = \Omega.$$

In other words, at c, Watson does not know that $\{c\}$ is true but he always knows that he doesn't know $\{c\}$. Hence $U_{MR}(\{c\}) = \emptyset$, so Modica-Rustichini would reject this possibility correspondence for failing to satisfy symmetry.

However, if one looks at the calculations given of $U(E)$ in Part B, one sees that it is symmetric! That is, the operator based on lack of knowledge at *every* level is symmetric, suggesting that Modica-Rustichini's rejection of this possibility correspondence may have been inappropriate. To understand this, let us reconsider the supposed unawareness of the event $\{a, b\}$. We saw that $\neg K \neg K(\{a, b\}) = \Omega$. Hence

$$K \neg K \neg K(\{a, b\}) = K(\Omega) = \Omega.$$

In other words, at state c, Watson does know that he doesn't know that he doesn't know $\{a, b\}$. This form of knowledge is certainly odd; however, it seems to suggest that he does indeed have positive knowledge of $\{a, b\}$ at some level and hence, we would argue, should *not* be said to be unaware of the possibility. Hence the fault seems to lie with declaring Watson to be unaware of this event, not the failure of symmetry.

Aside from the two papers by Modica and Rustichini, the only other papers we know of on unforeseen contingencies are Fagin and Halpern (1988), Kreps (1988), and

[5] Modica and Rustichini also use other assumptions that we will not need, in particular, positive introspection and nondelusion. While their proof does not use positive introspection in a critical manner, it does rely heavily on nondelusion.

Ghirardato (1995). Fagin and Halpern is the most closely related to our work in that they also discuss an unawareness operator. However, they do not develop the relationship with the possibility-correspondence literature and they do not consider the implications (such as the impossibility of nontrivial unawareness) of the kinds of properties we consider. Kreps and Ghirardato take decision-theoretic approaches. Kreps derives subjective states that represent unforeseen contingencies in the agent's mind. Ghirardato represents unforeseen contingencies by allowing acts which give sets of consequences as a function of the state. Neither relates unawareness to an information structure or gives an unawareness operator.

2. STATE-SPACE MODELS

Let Ω denote the state space. Any possibility correspondence P determines a knowledge operator K, mapping the power set of Ω into itself, as described above: $K(E)=\{\omega\in\Omega \mid P(\omega)\subseteq E\}$. Alternatively, one could simply begin with a knowledge operator. It turns out that only certain knowledge operators can be derived from an underlying possibility correspondence.[6] Therefore, starting from knowledge operators provides a strictly more general approach. Since our critique covers more than just possibility correspondence models, we require this extra generality.

The following properties of knowledge are usually assumed but will prove problematic for an agent who is unaware of something.

Necessitation $\quad K(\Omega)=\Omega,$

Monotonicity $\quad E\subseteq F\Rightarrow K(E)\subseteq K(F).$

Necessitation is the assumption that the agent "knows all tautologies." This name comes from the philosophy literature.[7] Monotonicity says that if E implies F, then knowledge of E implies knowledge of F.

The reader should suspect that there will be problems with making these assumptions hold in a model where the agent is unaware of some possibilities. Both seem to require the agent to have a certain understanding of the state space which seems questionable when the agent is unaware of something.

It is well-known that a knowledge operator is derivable from, and in fact equivalent to, a possibility correspondence if and only if it satisfies monotonicity, necessitation, and $K(E)\cap K(F)=K(E\cap F)$, an assumption we will never use.[8] The main result of this section is that knowledge operators that satisfy either necessitation or monotonicity cannot provide a nontrivial model of unawareness. In this sense, possibility correspondences preclude unawareness.

Analogously to the way knowledge is modeled, we suppose there is an operator $U:2^{\Omega}\rightarrow 2^{\Omega}$ with the interpretation that $U(E)$ is the set of states where the agent is unaware of the possibility that event E occurs. We do not construct this operator from the knowledge operator, but instead allow any operator satisfying certain axioms given below. We refer to the tuple (Ω, K, U) as a *standard state-space model*. Our use of the

[6] In fact, a knowledge operator, say K, that can be derived from a possibility correspondence, say P, is equivalent to P in the sense that we can uncover the original P by $P(\omega)=\cap\{E\subset\Omega \mid \omega\in KE\}$. See, e.g., Dekel and Gul (1997) for more detail.

[7] See Chellas (1980) for a discussion of this and other assumptions in the context of modal logic.

[8] See, for example, Dekel and Gul (1997). It is easy to show that this last assumption implies monotonicity.

phrase "standard" will be clarified much later when we give an example of a "nonstandard" state-space model.

We now consider three axioms on K and U that are the only assumptions we need for our first result. The first axiom says that the Modica-Rustichini definition is a necessary (though perhaps not sufficient) requirement for unawareness.

DEFINITION 1: The state-space model (Ω, K, U) is *plausible* if for every event $E, U(E) \subseteq \neg K(E) \cap \neg K \neg K(E)$.

Our perhaps overly strong terminology is based on the facts that the literature (often implicitly) views this as a necessary condition for unawareness and that it seems so obviously necessary to us—if the agent is unaware of an event, how can he know it is true or know that he doesn't know it?[9]

The next two axioms both require that this lack of positive knowledge extend to knowledge of unawareness itself.

DEFINITION 2: The state-space model (Ω, K, U) satisfies *KU introspection* if for every event E, $KU(E) = \emptyset$.

To see the intuition of this property, it is useful to divide it into two other properties. The first part is similar to the usual idea of nondelusion, the assumption that if the agent knows something (at least if it is something about his own knowledge), then it must be true. More specifically, the first part says that if the agent knows he is unaware, then he must be correct—that is, $KU(E) \subseteq U(E)$. The second part is analogous to the usual positive introspection, the assumption that if the agent knows something, then he knows that he knows it. Here the idea is that if the agent recognizes his own knowledge, then knowing he is unaware of a possibility should inform him about the possibility and hence make him aware of it. That is, $KU(E) \subseteq \neg U(E)$. Clearly, the only way for both of these properties to hold is if $KU(E) = \emptyset$. Less formally, the first property says that if an agent is aware of an event, he should not know falsely that he is unaware. The second says that if he is unaware of an event, his unawareness should preclude him from knowing this fact. Hence, either way, the agent cannot know that he is unaware of a specific event.

This seems consistent with our intuition about unawareness. While the agent may well know that he is unaware of *some* event, it doesn't seem reasonable to say that the agent knows precisely of which event he is unaware! The name we give this property is intended to emphasize the introspection part of the assumption which seems to us the stronger requirement.

The third axiom has a very similar intuition.

DEFINITION 3: The state-space model (Ω, K, U) satisfies *AU introspection* if $U(E) \subseteq UU(E)$.

[9] We emphasize that we use the word "unawareness" to mean that the agent fails to foresee a possibility, not that he knows of but does not understand some possibility. Plausibility seems completely reasonable in the former context, though not in the latter. To see the distinction, consider the statement "a doctor 50 years ago was unaware of the possibility of AIDS." By this, we mean that the doctor did not foresee the possibility of such a disease, not that if someone asked her what AIDS was, she would be unable to answer. In the latter case, she would surely know that she does not know what AIDS is.

AU introspection simply says that if an agent is unaware of an event E, then he must be unaware of being unaware. Another way to understand this property, which also explains its name, is to rewrite it in terms of awareness. So define $A(E) = \neg U(E)$. Then AU introspection is equivalent to $AU(E) \subseteq A(E)$. In other words, if the agent is aware of the possibility of being unaware of E, then he must be aware of E.

This idea is similar in spirit to KU introspection. Both properties essentially say that unawareness of an event precludes any form of positive knowledge regarding the event. As noted, Modica-Rustichini's definition of unawareness is stronger than plausibility. Also, they show that their assumptions imply KU and AU introspection. Hence they rely on strictly stronger hypotheses.

The following is the main result of this section.

THEOREM 1: *Assume (Ω, K, U) is plausible and satisfies both KU and AU introspection.*
(i) *If K satisfies necessitation, then for every event E, $U(E) = \varnothing$.*
(ii) *If K satisfies monotonicity, then for all events E and F, $U(E) \subseteq \neg K(F)$.*

In other words, necessitation implies that the agent is never unaware of anything. Monotonicity implies that in a state in which the agent is unaware of an event E, he cannot know *any* event F—that is, if the agent is unaware of anything, he knows nothing. Clearly, then, either property leaves us with only a trivial form of unawareness at best.

PROOF: By AU introspection and plausibility,

$$U(E) \subseteq U(U(E)) \subseteq \neg K \neg K(U(E)).$$

KU introspection is equivalent to $\neg KU(E) = \Omega$, so this implies $U(E) \subseteq \neg K(\Omega)$.

Since necessitation implies $K(\Omega) = \Omega$, it clearly implies $U(E) = \varnothing$ for all E. Monotonicity implies $K(F) \subseteq K(\Omega)$ for all F, so $\neg K(\Omega) \subseteq \neg K(F)$ for all F. Hence monotonicity implies $U(E) \subseteq \neg K(F)$ for all E and F. *Q.E.D.*

In short, a nontrivial model of unawareness requires us to abandon both necessitation and monotonicity. If we drop even one of these properties, we are restricted to knowledge operators that cannot be derived from possibility correspondences; hence nonpartitional possibility correspondences cannot appropriately model unawareness.[10]

If we do abandon necessitation and monotonicity, can we find standard state-space models which satisfy our three axioms in a nontrivial way? The following example shows that this is possible.

[10] One may wonder which of our axioms we could retain in a possibility correspondence model. We view plausibility as the most basic of the three axioms, so it seems natural to only consider relaxations of our assumptions which keep this. Tim van Zandt (private communication) has generalized a result in an earlier draft of this paper to show that any knowledge operator satisfying plausibility, monotonicity, and weak nondelusion (that is, $KK(E) \subseteq K(E)$ and $K \neg K(E) \subseteq \neg K(E)$) must satisfy KU introspection. The proof is simple and is contained in the working paper version of the present paper. The example in the introduction comes from a possibility correspondence and satisfies weak nondelusion, plausibility, and KU introspection with nontrivial unawareness. Hence we see that the primary conflict is between AU introspection and possibility correspondences.

EXAMPLE 2:

The knowledge operator is similar to the one used in Example 1. Again, let $\Omega = \{a, b, c\}$ and let K be given by $K(\emptyset) = \emptyset$, $K(\Omega) = \{a, b\}$, $K(\{a\}) = \{a\}$, $K(\{b\}) = \{b\}$, $K(\{c\}) = \emptyset$, $K(\{a, b\}) = \{a, b\}$, $K(\{a, c\}) = \{a, c\}$, and $K(\{b, c\}) = \{b\}$. Finally, we set $U(\{c\}) = U(\{a, b\}) = \{c\}$ and $U(E) = \emptyset$ for every other event E.[11]

This definition has the property that

$$UE \subseteq \bigcap_{i=1}^{\infty} (\neg K)^i E,$$

as can be easily verified. Hence this state-space model is plausible. It is not hard to verify that this model also satisfies KU and AU introspection as well as symmetry. Furthermore, at c, the agent does have some positive knowledge since he knows $\{a, c\}$ despite the fact that he is unaware of $\{c\}$ and $\{a, b\}$.

In light of Example 2, we see that it is possible to generate nontrivial unawareness from a standard state-space model, even when the three axioms above are satisfied. Does this example then imply that standard state-space models can be used to study unawareness?

3. PROPOSITIONAL MODELS

It is difficult to evaluate examples like Example 2 in the absence of some more concrete notion of what generates the model. One way to think about a given state-space model is that it arises from what we will call a *propositional model*.[12] In this section, we briefly describe such models, show how state-space models can be derived from them, and use this framework to suggest another criterion that a reasonable state-space model should satisfy. We then show that this criterion rules out Example 2 and, in fact, rules out nontrivial unawareness in all such state-space models.

A propositional model explicitly introduces the statements about the world that the agent might know or be aware of. To give the simplest possible version of such a model, we avoid giving a full blown modal logic, but instead give a small part of a fuller logic. More specifically, assume we have a set Φ which we refer to as the set of *propositions* (or, more properly, propositional formulae). An element of this set should be interpreted as a statement about the world which might be true or false. We assume that this set is *closed* in the sense that if $\varphi \in \Phi$, then there are other propositions in Φ denoted $u\varphi$, $k\varphi$, and $\neg\varphi$, as well as a special proposition we denote $\top(\varphi)$. The first three are respectively interpreted as the statement that the agent is unaware of φ, the agent knows φ, and φ is false. The proposition $\top(\varphi)$ is interpreted as an "obvious" tautology involving φ and no other proposition. For example, we could think of $\top(\varphi)$ as the statement "φ implies φ" or "φ is true if and only if φ is true" or "either φ is true or it is false."

Second, we introduce a function $T : \Phi \to 2^{\Omega}$ where $T(\varphi)$ is the set of states in which φ is true. We refer to (Φ, Ω, T) as a *propositional model*.

[11] The knowledge operator differs from Example 1 only in $K(\{a, c\})$, which there equalled $\{a\}$.

[12] A propositional model, like a Kripke model (see, e.g., Fagin et al. (1995), is a semantic model in which the connection to the syntax is made explicit.

It is not hard to show that we can recover standard state-space models from propositional models by imposing two assumptions. The first assumption is that all states are *real*. That is, in any state of the world each proposition is either true or false but not both, and the "obvious" tautologies are always true.

DEFINITION 4: A propositional model (Φ, Ω, T) has *real states* if

$$T(\varphi) = \Omega \setminus T(\neg \varphi)$$

and

$$T(\top(\varphi)) = \Omega.$$

The second assumption needed to recover a standard state-space model is that the only aspect of a proposition relevant for determining whether the agent knows or is aware of it is the subset of the state space in which it holds.

DEFINITION 5: A propositional model (Φ, Ω, T) satisfies *event sufficiency* if

$$T(\varphi) = T(\psi) \Rightarrow T(k\varphi) = T(k\psi) \quad \text{and} \quad T(u\varphi) = T(u\psi).$$

In other words, event sufficiency says that if two propositions correspond to the same event in the sense that they are true in exactly the same subset of Ω, then the agent knows (is aware of) one if and only if he knows (is aware of) the other.[13]

DEFINITION 6: A propositional model (Φ, Ω, T) is *standard* if it satisfies real states and event sufficiency.

It is easy to see that a standard propositional model can always be recast as a standard state-space model. In particular, we can define

$$K(E) = T(k\varphi) \quad \text{for any } \varphi \text{ such that } T(\varphi) = E$$

and

$$U(E) = T(u\varphi) \quad \text{for any } \varphi \text{ such that } T(\varphi) = E.$$

In words, $K(E)$ is simply the set of states where $k\varphi$ is true for any φ with $T(\varphi) = E$ and analogously for $U(E)$. Event sufficiency implies that the particular φ chosen is irrelevant. Hence event sufficiency means that we can identify the set of states where the agent knows or is aware of something without reference to the underlying propositions. Similarly, the real states assumption means that if we wish to identify the set of states where the agent does *not* know or is *not* unaware of something, we can simply take the complement of the event where he does know or is unaware of this statement. In short, if we are only interested in knowledge and unawareness in the abstract and not the specific statements involved, we can recast a standard propositional model into a standard state-space model without losing any of the information of interest. In this sense, we can think of the class of standard propositional models as being the "foundation" for standard state-space models.

[13] This condition appears as LE in Fagin et al. (1995, p. 318) and as RE in Lismont and Mongin (1994).

A natural requirement for a propositional model is that if the agent is aware of φ, then he knows any "obvious" tautology involving φ. If he is aware of φ, surely whatever unawareness he has about other possibilities does not confuse his ability to recognize that, say, "φ implies φ" must be true. That is, it seems natural to require

$$T(\neg u\varphi)\subseteq T(k\top(\varphi)).$$

It is easy to see that if we require this in a standard propositional model, then the corresponding state-space model will satisfy

$$\neg U(E)\subseteq K(\Omega).$$

This is the translation of the property above since the event in which $\top(\varphi)$ holds is Ω by the real-states assumption. We call this property *weak necessitation* since it is an obvious weakening of necessitation.

It is easy to see that Example 2 does not satisfy weak necessitation since $\neg U(\Omega)=\Omega$ but $K(\Omega)=\{a,b\}$. We now show that only trivial unawareness is possible in any standard state-space model satisfying our three axioms plus weak necessitation.

THEOREM 2: *Given any standard state-space model satisfying plausibility, KU and AU introspection, and weak necessitation, for all events E, F, and G,* $U(E)=U(F)\subseteq\neg K(G)$.

In other words, if the agent is unaware of anything, he is unaware of everything and knows nothing.

PROOF: As shown in the proof of Theorem 1, the three axioms imply $U(E)\subseteq\neg K(\Omega)$. Weak necessitation implies $\neg K(\Omega)\subseteq U(F)$ for all F. Hence we obtain $U(E)\subseteq U(F)$ for all E and F, so $U(E)=U(F)$ for all E and F. Obviously, plausibility implies $U(G)\subseteq\neg K(G)$, so we get $U(E)=U(F)\subseteq\neg K(G)$ for all events E, F, and G. *Q.E.D.*

Another advantage of explicitly introducing standard propositional models is that they enable us to clarify the nature of Theorems 1 and 2. The key to both theorems is that we use plausibility and AU introspection to identify an event the agent cannot know if he is unaware of E. Specifically, these axioms imply that if the agent is unaware of E, then he cannot know the event $\neg KU(E)$. KU introspection says that this event is, in fact, Ω. Hence we get $U(E)\subseteq\neg K(\Omega)$. We then show that adding any of necessitation, monotonicity, or weak necessitation eliminates nontrivial unawareness.

A natural question to ask in light of this is whether we should simply assume $U(E)\subseteq\neg K(\Omega)$ by assuming $T(u\varphi)\subseteq T(\neg k\top(\varphi))$ and event sufficiency. In other words, if the agent is unaware of φ, shouldn't it be true that he doesn't know any statement involving φ, whether it is a tautology or not? We do not find such an assumption compelling since it is quite difficult to understand what is means for an agent to know or not know a tautology which involves something of which he is unaware. Does he recognize a statement like "φ implies φ" as obviously true even if he does not have a clue what φ is? Or does the fact that he has not thought of the possibility of φ mean that this statement is simply not in his internal language? Our proof does not require us to take a stand on this philosophically troubling point. We *derive* the lack of knowledge from principles we believe to be necessary aspects of unawareness, rather than simply assuming this lack of knowledge. It is also important to note that weak necessitation does not require a position on this difficult point. Weak necessitation refers to the implications of being aware, not unaware. Surely if the agent is aware of φ, he has no problems recognizing that "φ implies φ" must be true.

4. CONCLUSION

Propositional models are useful for a third reason: they help identify why standard state-space models are unable to capture a nontrivial notion of unawareness. Clearly, it is the combination of the real-states and event-sufficiency assumptions which causes the problem. While event sufficiency may appear less intuitive and hence more culpable than real states, there is a very simple reason why it is at least as natural to drop the real-states assumption. In standard state-space models, states play two distinct roles: they are the analyst's descriptions of ways the world might be and they are also the agent's descriptions of ways the world might be. If the agent is unaware of some possibility, though, "his" states should be less complete than the analyst's. In particular, the propositions the agent is unaware of should not "appear in" the states he perceives. In other words, if the agent is unaware of φ, then φ should be neither true nor false in the states he considers possible. That is, the states he considers possible are in neither $T(\varphi)$ nor $T(\neg \varphi)$, violating $T(\varphi) = \Omega \setminus T(\neg \varphi)$.

Once we think of the state space as representing the agent's view of possibilities, event sufficiency—and, for that matter, necessitation and monotonicity—become much more intuitively appealing. If we think of $T(\varphi) = T(\psi)$ as meaning that *in the agent's view*, φ and ψ are equivalent, then it seems quite reasonable to say that he should know one is true if and only if he knows the other is true—that is, $T(k\varphi) = T(k\psi)$. Similarly, if $T(\varphi) = \Omega$, the subjective view of states would lead us to interpret this as saying that the agent always believes that φ is always true. Hence it seems natural to suppose that $T(k\varphi) = \Omega$, precisely what necessitation says in this framework.

This suggests that we enrich the standard state-space model by allowing for subjective states in which some proposition may be neither true nor false. If we do, a proposition will be described not by a set of states where it is true, but by a set where it is true and another set where it is false. (With real states, by definition, the second set is the complement of the first.) This requires us to consider knowledge and awareness operators that are also richer. Instead of considering $K(E)$ (where E is the set of states where some proposition is true), we are led to consider $K(E, F)$, where the way (E, F) corresponds to a proposition φ is that E is the set of states where φ is true and F the set where φ is false. Similarly, since $K(E, F)$ itself should correspond to a proposition—namely, the proposition that the agent knows the proposition which corresponds to (E, F)—it will also consist of two sets. That is, $K(E, F)$ will be written as $(K_t(E, F), K_f(E, F))$, where $K_t(E, F)$ is the set of states where it is true that the person knows (E, F) and $K_f(E, F)$ is where it is false. In the Appendix, we describe this structure, which is based loosely on Modica-Rustichini (1993), in the context of an example.

As we show in the example, this kind of nonstandard state-space model is not subject to the problems noted in Theorems 1 and 2. Hence we conclude that our critique of state-space models is indeed restricted to the class of standard state-space models. We do not claim that the approach used in the example necessarily provides a useful model of unawareness. We view the axioms we have given as necessary, not sufficient, conditions for an interesting model of unawareness. Whether or not this approach is useful hinges on whether it can provide the epistemic basis for an interesting and useful decision theory. The exploration of these possibilities is beyond the scope of this paper.

Dept. of Economics, 2003 Sheridan Rd., Northwestern University, Evanston, IL 60091, U.S.A.; dekel@nwu.edu,

Dept. of Economics, Social Science Centre, University of Western Ontario, London, Ontario N6A 5C2, Canada; blipman@julian.uwo.ca,

CentER, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; aldo@kub.nl.

Manuscript received February, 1996; final revision received February, 1997.

APPENDIX

As noted, we need two different kinds of states—the *real states* which give full and complete descriptions of ways the world might be and the *subjective states*, the descriptions of possibilities as perceived by the agent. Let Ω_R denote the real states and Ω_S the subjective states (where these sets are disjoint).

Any given real state, $\omega \in \Omega_R$, will include a description of how the agent views the world. Hence it will include a specification of which subjective states the agent thinks are possible. Formally, we represent this in two steps. First, we have a function $\rho : \Omega_R \to \Omega_S$ with the interpretation that $\rho(\omega)$ is the subjective state the agent is "in" when the real state is ω. That is, $\rho(\omega)$ is ω but filtered through the agent's eyes. Second, we have a partition, Π, of the subjective states. For any $\omega \in \Omega_S$, let $\pi(\omega)$ denote the event in Π containing ω. The interpretation is that in real state ω, the set of subjective states the agent views as possible are those in $\pi(\rho(\omega))$.

As discussed in the text, we identify each proposition with a pair of events, rather than a single event. Not every pair of events will necessarily correspond to a proposition. For example, there is no obvious reason why we should allow a proposition to be both true and false, suggesting that we should restrict attention to event pairs which are disjoint. A pair of events which does correspond to a proposition will be called a *statement*. We let S denote the set of statements. Since we only give an example here, we avoid a detailed construction and simply assume that if $(E, F) \in S$, then E and F are disjoint. We also suppose that if $(E, F) \in S$, then $(F, E) \in S$. Note that (F, E) is naturally thought of as the negation of (E, F) since (F, E) is true precisely when (E, F) is false, and (F, E) is false when (E, F) is true.

The knowledge and unawareness operators will map statements into statements. That is, $K(E, F)$ is interpreted as the statement that the agent knows the statement (E, F). As noted in the text, $K(E,F) = (K_t(E, F), K_f(E, F))$, where $K_t(E, F)$ is the set of states where it is true that the agent knows (E, F) and K_f is the set of states where it is false. Similarly, the unawareness operator is written $(U(E, F), A(E, F))$ where $U(E, F)$ is the set of states where it is true that the agent is unaware of (E, F) and $A(E, F)$ is the set where it is false that he is unaware of (E, F).

To be more concrete, suppose there are two states, a and b, where a is real and b is subjective. Of course, $\rho(a) = b$ and the partition Π consists of a single event containing the single state b. So in a, the agent believes that b is the true state of the world. Consider any statement which is true in state b. Because the agent believes b is the true state, the agent "knows" this statement to be true. Similarly, any statement which is false in state b is one that the agent does not know to be true. Since a is the real state and is the complete description corresponding to the incomplete description b, these statements are also true at a. More precisely,

$$(K_t(E, F), K_f(E, F)) = (\{a, b\}, \varnothing) \quad \text{if } b \in F$$

and

$$(K_t(E, F), K_f(E, F)) = (\varnothing, \{a, b\}) \quad \text{if } b \in E.$$

Given our restriction on statements, we can never have $b \in E$ and $b \in F$. Hence the only remaining case is where $b \notin E$ and $b \notin F$. In this case, the agent is unaware of (E, F). In the real state, a, an omniscient outside observer would certainly say that the agent does not know (E, F) to be true. On the other hand, the agent doesn't recognize this possibility so, at b, it is neither true nor false that he knows (E, F). Hence we define

$$(K_t(E,F), K_f(E,F)) = (\varnothing, \{a\}) \quad \text{if } b \notin E \text{ and } b \notin F.$$

The unawareness operator is similarly defined. If $b \in E$ or $b \in F$, then the agent is well aware of (E, F). It is only when b is not in either event that the agent is unaware of (E, F). Hence

$$(U(E,F), A(E,F)) = \begin{cases} (\varnothing, \{a,b\}) & \text{if } b \in E \cup F, \\ (\{a\}, \varnothing) & \text{otherwise}. \end{cases}$$

We now show that natural analogs of our three axioms are satisfied. First, note that $U(E,F)$ is nonempty only when $b \notin E \cup F$ and equals $\{a\}$ in this situation. In this same case,

$$K(E,F) = (K_t(E,F), K_f(E,F)) = (\varnothing, \{a\}),$$

so

$$K(K_f(E,F), K_t(E,F)) = K(\{a\}, \varnothing) = (\varnothing, \{a\}),$$

and likewise for higher levels. In other words, when the agent is unaware of (E, F), it is also true that he does not know (E, F), does not know that he does not know it, etc., for all possible iterations. Hence plausibility is satisfied. Also,

$$U(E,F) = \begin{cases} \varnothing = U(U(E,F), A(E,F)) & \text{if } b \in E \cup F, \\ \{a\} = U(U(E,F), A(E,F)) & \text{if } b \notin E \cup F. \end{cases}$$

In other words, if the agent is unaware of (E, F), then he is unaware that he is unaware, so *AU* introspection holds. Finally,

$$K(U(E,F), A(E,F)) = \begin{cases} K(\varnothing, \{a,b\}) = (\varnothing, \{a,b\}) & \text{if } b \in E \cup F, \\ K(\{a\}, \varnothing) = (\varnothing, \{a\}) & \text{if } b \notin E \cup F. \end{cases}$$

That is, the set of states at which the agent knows he is unaware of (E, F) is empty. So *KU* introspection holds.

Finally, we show that natural analogs of necessitation, monotonicity, and weak necessitation hold as well. First, note that

$$K(\{a,b\}, \varnothing) = (\{a,b\}, \varnothing),$$

so that it is always true that the agent knows $(\{a,b\}, \varnothing)$. Hence a natural analog of necessitation holds, so the same is true of weak necessitation. Second, it is easy to see that if $E \subseteq E'$, then $K_t(E,F) \subseteq K_t(E',F)$, a natural analog of monotonicity. Hence one can satisfy (analogs of) necessitation, weak necessitation, and monotonicity along with our three axioms in a nontrivial but nonstandard state-space model. It is not hard to show that these results hold in general for these models.

REFERENCES

BRANDENBURGER, A., E. DEKEL, AND J. GEANAKOPLOS (1992): "Correlated Equilibrium with Generalized Information Structures," *Games and Economic Behavior*, 4, 182–201.

CHELLAS, B. (1980): *Modal Logic: An Introduction.* Cambridge: Cambridge University Press.

DEKEL, E., AND F. GUL (1997): "Rationality and Knowledge in Game Theory," in *Advances in Economic Theory, Seventh World Congress*, ed. by D. M. Kreps and K. F. Wallis. Cambridge: Cambridge University Press, pp. 87–172.

FAGIN, R., AND J. Y. HALPERN (1988): "Belief, Awareness and Limited Reasoning," *Artificial Intelligence*, 34, 39–76.

FAGIN, R., J. Y. HALPERN, Y. MOSES, AND M. Y. VARDI (1995): *Reasoning About Knowledge*. Cambridge: MIT Press.

GEANAKOPLOS, J. (1989): "Game Theory without Partitions, and Applications to Speculation and Consensus," Yale University Working Paper.

GHIRARDATO, P. (1995): "Coping with Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty," Cal Tech Working Paper.

HART, O. (1995): *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press.

KREPS, D. (1988): "Static Choice in the Presence of Unforeseen Contingencies," Stanford University Working Paper.

LISMONT, L., AND P. MONGIN (1994): "A Non-minimal but Very Weak Axiomatization of Common Belief," *Artificial Intelligence*, 70, 363–374.

MODICA, S., AND A. RUSTICHINI (1993): "Unawareness: A Formal Theory of Unforeseen Contingencies: Part II," CORE Discussion Paper.

——— (1994): "Awareness and Partitional Information Structures," *Theory and Decision*, 37, 107–124.

MORRIS, S. (1996): "The Logic of Belief and Belief Change: A Decision Theoretic Approach," *Journal of Economic Theory*, 69, 1–23.

MORRIS, S., AND H. SHIN (1992): "Noisy Bayes Updating and the Value of Information," University of Pennsylvania Working Paper.

SAMET, D. (1990): "Ignoring Ignorance and Agreeing to Disagree," *Journal of Economic Theory*, 52, 190–207.

SHIN, H. (1993): "Logical Structure of Common Knowledge," *Journal of Economic Theory*, 60, 1–13.