Minnesota Economics Research Reports

Object-Based Unawareness

by

Oliver Board
University of Pittsburgh

and

Kim-Sau Chung
University of Minnesota

Research Report No. 2007-02, September 2007

Department of Economics

UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455

Object-Based Unawareness

Oliver Board and Kim-Sau Chung
University of Pittsburgh University of Minnesota

August 24, 2007

Abstract

The goal of this paper is to construct a user-friendly model of unawareness. We start from an axiom system (in first-order modal logic) that enables us to express the following kinds of sentences: "the agent is not sure whether or not there is anything that he is unaware of", and "I am not sure whether or not you are aware of something that I am not". We then prove a characterization theorem which describes a class of structures, called object-based unawareness structures, that correspond to this axiom system. As an application, we explain how an object-based unawareness structure can be use to model those American founding fathers who were opposed to the inclusion of the Bill of Rights in the constitution.

1 Introduction

It has been recognized for some time that unawareness is an important phenomenon in economic life. Consider, for example, contracting parties faced with a vast number of unknown unknowns. They may choose to write incomplete contracts in order to retain the flexibility to deal with contingencies they were unaware of at the time of agreement. Similarly, if legislators are unaware of every potential bad action, they may choose to write vague laws for fear of how judges will interpret (inevitable) omissions from more precise ones. Commenting on whether the Bill of Rights should be included in the Constitution of the United States, James Iredell (later a Supreme Court Justice) made the following remarks: "[it would be] not only useless, but *dangerous*, to enumerate a number of rights which are not intended to be given up; because it would be implying, in the strongest manner, that every right not included in the exception might be impaired by the government without usurpation." Chung and Fortnow (2007) develop a theory of legislation along these lines. They argue that legislators who are not sure whether or not there are any bad actions that they are unaware of may optimally refrain from enumerating all the bad actions that they are aware of.

We can call sentences such as "the legislators are not sure whether or not there are any bad actions that they are unaware of" or "I am not sure whether or not you are aware of something that I am not" *EU-sentences*, where EU stands for "reasoning about the *E*xistence of things that one is *U*naware of".[1] Any model, including Chung and Fortnow (2007), that tries to model agents such as James Iredell has first to express EU sentences. However, RU sentences are not easy to express in traditional economic models such as those based on information partitions. In fact, such models preclude the expression of an even simpler class of sentences, which we shall call *U-sentences*, such as: "an agent is aware of this but not aware of that." (We shall explain the expressibility problem in more detail below.) The observation that these models preclude the expression of U-sentences is formalized in a seminal paper by Dekel, Lipman, and Rustichini (1998) (hereafter DLR) titled "Standard State-Space Models Preclude Unawareness." EU-sentences are more complicated than U-sentences in that the former involve existential quantifiers which allow us to describe unspecified objects, while the latter do not.

There have been a number of recent studies that construct non-standard models that do not preclude the expression of U-sentences. Examples include Modica and Rustichini (1999), Halpern (2001), Feinberg (2004), Li (2006), and Heifetz, Meier, and Schipper (2006). But the problem of expressing EU-sentences remains open. The goal of this paper is to construct a user-friendly model that can express these sentences. We start from an axiom system that does not preclude the expression of EU-sentences. We then prove a characterization theorem which describes the class of structures that correspond to this axiom system. We call this the class of structures the *object-based unawareness structures*. As an application, we explain how one can use an object-based unawareness structure to model agents such as James Iredell.

It goes without saying that we owe huge intellectual debt to each of the unawareness papers cited above. Without this earlier literature, the current paper would not be possible. A brief review of this literature will be provided in Section 6.

Two recent independent studies, Halpern and Rego (2006) and Sillari (2006), also address the issue of EU-sentences. They employ a different way of dealing with the existential quantifier, adopting the possibilitist (rather than the actualist) approach, which leads them to construct structures with constant domains (in contrast, our structures have non-constant domains, which we shall explain faciliate the modelling of an agent who is not sure whether or not there exist things he is unaware of); and, more importantly, they model awareness in a very different way, taking what we call a "semi-syntactic" approach. We shall review their approaches in Section 6 as well.

1.1 The Expressibility Problem

In this section, we shall briefly review two questions: (1) In what sense are U-sentences not expressible in standard state-space models, such as models with partitional information structure? (2) In what sense are EU-sentences not expressible in non-standard state-space models, such as the models developed in the earlier unawareness literature? Even if the reader

[1] In this informal section, we use "I am not sure whether or not ϕ" as a shorthand for the conjunction of two sentences: "I do not know that ϕ is true" and "I do not know that ϕ is not true".

is already familiar with this literature, we still urge that he skims through this section, as we shall introduce some notation and terminology here.

First, we note that DLR do not commit to a specific definition of unawareness. Rather, their impossibility theorem applies to all definitions that are "reasonable" in a sense they make precise, and can be expressed in the context of a standard state-space model. For DLR, a standard state-space model is a state space $\mathcal{W}$ (a set of states) coupled with a knowledge operator k, which is a mapping from events to events (recall that an event is a subset of the state space), with the requirement that either $k(\mathcal{W}) = \mathcal{W}$ (Necessitation) or $k(E) \subset k(F)$ as long as $E \subset F$ (Monotonicity). Within these models, they consider definitions of unawareness that take the form of an operator (again, a mapping from events to events). They argue that any sensible definition of unawareness should satisfy three axioms:[2]

1. $u(E) \subset (\mathcal{W} \setminus k(E)) \cup (\mathcal{W} \setminus k(\mathcal{W} \setminus k(E)))$ (Plausibility),
2. $U(E) \subset U(U(E))$ (AU Instrospection), and
3. $K(U(E)) = \emptyset$ (KU Introspection).

They prove that if the unawareness operator satisfies these three axioms, then necessitation implies that the agent is aware of everything, and monotonicity implies that if the agent is unaware of anything, he knows nothing. In standard state-space models, then, the three axioms are inconsistent with a nontrivial notion of unawareness.

At the risk of being repetitive, let's illustrate this expressiblity problem with an example. During this illustration we shall also introduce some new notation and terminology, which will eventually allow us to restate DLR's impossibility result in a slightly different way.

Consider the following state space: $\mathcal{W} = \{w_1, w_2\}$. In state w_1, the agent cannot distinguish w_1 and w_2, and consider both possible. In state w_2, the agent consider only w_2 possible. According to the usual definition in this framework, an agent knows an event if and only if it holds at every state he considers possible, so the corresponding knowledge operator is such that $k(\mathcal{W}) = \mathcal{W}$, $k(w_1) = k(\emptyset) = \emptyset$, and $k(w_2) = w_2$.[3]

As an application, suppose fact ϕ is true in and only in state w_2, and let's write this as follows:

$$(M, w_1) \models \neg\phi,$$
$$(M, w_2) \models \phi,$$

where M refers to the model we outlined in the paragraph above, and $\models$ represents truth (so $(M, w_1) \models \neg\phi$ means "not ϕ is true in world 1 of model M).

[2] In words, these axioms say that "if an agent is unaware of something, then he does not know it, and he does not know that he does not know it", "if an agent is unaware of something, then he must be unaware that he is unaware of it", and "an agent cannot know that he is unaware of a specific fact".

[3] We ease notation by omitting brackets when a set is a singleton.

In this model M, because $k(w_2) = w_2$, if we use $K_i\phi$ to stand for "the agent knows fact ϕ", then $K_i\phi$ is true in and only in w_2. Using the notation just introduced, we have:

$$(M, w_1) \models \neg K_i\phi,$$
$$(M, w_2) \models K_i\phi.$$

Similarly, because $k(w_1) = \emptyset$, we have:

$$(M, w_1) \models \neg K_i\neg K_i\phi,$$
$$(M, w_2) \models \neg K_i\neg K_i\phi.$$

What we are doing above is assigning truth values to longer and longer sentences such as ϕ, $\neg\phi$, $K_i\phi$, $\neg K_i\phi$, and $\neg K_i\neg K_i\phi$. How about sentences such as "the agent is unaware that it is raining"? How should we assign a truth value to $U_i\phi$? This is a new kind of sentence, involving the letter U_i instead of (merely) K_i. It is an example of what we have called a U-sentence. Defining unawareness in model M amounts to extending the truth-value assignment rule to U-sentences.

One natural attempt is to extend the truth-value assignment rule as follows: whenever a sentence of the form $\neg K_i\neg K_i\alpha$ is true in a state w, let's also assign "true" to the sentence $U_i\alpha$ in that same state. Applying this rule to our example, we have:

$$(M, w_1) \models U_i\phi,$$
$$(M, w_2) \models U_i\phi.$$

But then, since $k(\mathcal{W})$, we have to also assign "true" to the sentence $K_iU_i\phi$ in both states, an assignment that violates DLR's KU Introspection Axiom.

How about other ways of extending the truth-value asignment rule? One way to paraphrase DLR's impossiblity result is as follows: in any standard state-space model M, there is no way to extend the truth-value assignment rule to U-sentences without violating at least one of DLR's three axioms (modulo some trivial extensions).

We are now in position to address the second question: In what senses are EU-sentences not expressible in those non-standard state-space models introduced in the earlier unawareness literature? Technically, it is incorrect to say that EU-sentences are inexpressible in those models. In each of those models, there is a corresponding truth-value assignment rule, which assigns truth values to U-sentences in a way that does not violate any of DLR's three axioms. But these rules do not specify how truth values should be assigned to EU-sentences, nor has anyone proposed any extension of these rules to EU-sentences. Indeed, no one has even suggested any axioms, analogous to DLR's three axioms, that any sensible extension should satisfy.

In this paper, instead of working with the existing non-standard state-space models and trying to extend their truth-value assignment rules to EU-sentences, we shall start from scratch. We shall start with an explicit list of "reasonable" axioms, analogous to DLR's axioms, and then construct a class of structures (together with a truth-value-assignment rule for a rich enough set of sentences that includes EU-sentences) that is axiomatized by exactly those axioms. The advantage of starting from scratch is apparent: since no axioms will be hidden from the readers, we can afford a more open discussion of whether these structures (as well as the corresponding truth-value assignment rule) are sensible or not.

1.2 The Main Idea

In this section, we shall very informally explain the main idea that underlies our axioms (and hence also our structures). This idea can be summarized as follows: while knowledge is often associated with events/facts, unawareness is more often assocated with objects/things.

This difference between knowledge and unawareness has actually already appeared in many motivating examples provided by previous authors. For instances, Heifetz, Meier, and Schipper (2006) write "Scientists were unaware of gravity until Newton conceived it. Mathematicians are now unaware of tomorrow's proof-techniques for today's long-standing conjectures. Some investors are unaware of financial market regularities that other investors exploit". Note that "gravity", "techniques", and "regularities" are more akin to objects rather than events. Even when we sometimes say that an agent is unaware of a certain fact, as in a sentence like "I was not aware that people could use this loophole to circumvent the law", the agent's unawareness of a certain fact can often be traced to his unawareness of a certain object referred to in the fact (here, the object is the loophole in question). The reason behind this unawareness may be very mundane (e.g., the loophole in question simply slipped the mind of the busy legislator when he wrote the law), or quite fundamental (e.g., before viruses were discovered, the word "virus" amounted to nothing more than a funny sounding word to a typical ear), but in both cases the effect on the agent's knowledge is the same: if an agent is not even aware of every object referred to in a particular fact, then he cannot know this fact, and he cannot know that he does not know this fact . . . , etc.

However, while a legislator who is unaware of loophole x certainly cannot know that "people can use loophole x to circumvent law y" (nor can he know that he does not know this), he may well know that "there exist some loopholes that people can use to circumvent law y", because the latter fact does not mention x, the object that he is unaware of. In other words, an agent can be aware of his own unawareness of "something", yet not know what that "something" is.

Our axioms will take this idea seriously: in effect they say that an agent's unawareness of a fact can be traced to his unawareness of the mentioned objects, and if he is aware of all objects mentioned then he is also aware of that fact. It could be argued that we have taken this idea too far, with the result that we fail to capture alternative uses to the word "aware" in everyday English. For example, in a sentence like "I was not aware that you were in town, otherwise I would have bought you dinner", the agent's unawareness of a fact ("that you were in town") is definitely not due to his unawareness of objects. In this example, however, “aware” is synonymous with “know”, and the same meaning could have been expressed without using the word “aware”: “I did not know that you were in town, otherwise I would have bought you dinner”. In such cases, then, “aware” is not being used to describe a new concept, and we can ignore this alternative use without loss.[4]

1.3 Structure of the Paper

Section 2 contains our main result. Section 2.1 first defines the language, or equivalently, the set of sentences that we are to assign truth values to. Section 2.2 presents our axioms. Section 2.3 then presents the class of structures, which we shall call the object-based unawareness (OBU) structures, that is axiomatized by exactly the axioms presented in Section 2.2. Section 2.4 formally states our characterization theorem.

Section 3 gives an example of our framework in use, showing how one can use an OBU structure to model those American founding fathers who were opposed to including the Bill of Rights in the constitution.

In section 4, we verify that our structures satisfy DLR’s three axioms. Then, in Section 5, we discuss how to incorporate other axioms of interest.

Section 6 reviews the previous unawareness literature, and Section 7 concludes.

2 Object-Based Unawareness

In this section, we shall first introduce our language. A *language* can be thought of as the set of all sentences, or more precisely all sentences that we would like to assign truth values to in the structures that we are about to construct in Section 2.3. Our language will be based on first order modal logic, and is different from the language used in the previous unawareness literature, which in turn is based on propositional modal logic. Introducing this new language is necessary, because EU-sentences cannot be expressed in propositional modal logic.

Although we do not describe object-based unawareness (OBU) structures until Section 2.3, let’s use $\mathcal{M}$ to denote the class of such structures that characterizes the axiom system we are about to introduce, with M being a typical structure within that class. Within the set of all sentences, there is a subset that is of particular interest, namely those sentences that are *valid* in $\mathcal{M}$. the definition of a valid sentence will also have to wait until Section 2.3, but roughly speaking valid sentences are those sentences that are always true in every structure M in $\mathcal{M}$. They are intimately related to theorems that economists may one day

[4] We should also note that the word “know” has also occasionally been used in a way that is similar to how we use the word “aware” here. A famous example is Russell (1912): “The word ‘know’ is here used in two different senses. (I) In its first use it is applicable to the sort of knowledge which is opposed to error, the sense in which what we know is *true*, the sense which applies to our beliefs and convictions, i.e. to what are called *judgements.* [...] This sort of knowledge may be described as knowledge of *truths.* (2) In the second use of the word ‘know’ above, the word applies to our knowledge of *things*, which we may call acquaintance.”

prove using OBU structures. In fact, the notion of validity (if not the word itself) appears in other contexts that will be familiar to many economists. For example, the sentence “it cannot be common knowledge that agents disagree on the probability of the same fact” is a valid sentence in the class of partitional information structures with a common prior, and is a result proved by Aumann (1976) using those structures. Axiomatization of a class of structures such as $\mathcal{M}$ is in effect axiomatization of its set of valid sentences.

Axiomatization of the set of valid sentences takes the form of a procedure to generate a set of *provable formulas* that coincides with the set of valid sentences. The procedure has two parts. First, some sentences are declared provable directly. These sentences are called *axioms*. Second, sentences other than axioms can also qualified as provable indirectly by association with existing provable formulas. The association rules are called *inference rules*. One way to interpret this exercise is to think of axioms and inference rules as two different forms of “hidden assumptions” of the class of structures being axiomatized. Every provable formula one may prove using a given class of structures must have its root in some of these “hidden assumptions.”

One last remark before we start: although we have been using the word “sentences” casually so far, we shall stop doing so below. The reason is that logicians typically reserve the word “sentences” for something else, and use the word *formulas* to refer to what we have been calling “sentences.” We shall follow this convention below.

2.1 The Language

Our language (to be formally defined shortly) is a version of first order modal logic. Roughly speaking, first order modal logic is first order logic augmented by modal operators, and first order logic is an extension of propositional logic. Examples of formulas in propositional logic include $\neg\alpha$ (read “it is not the case that α”), $\alpha \wedge \beta$ (read “α and fact β”), $\alpha \rightarrow \beta$ (read “whenever α is true, β will be true as well”), etc. First order logic extends propositional logic by including formulas such as $\forall x\, Tall(x)$ (“every x is tall”). Modal operators are represented by letters such as K_i and A_i that can be affixed to simpler formulas and result in longer ones: $K_i\alpha$ (read “agent i knows that α”) and $A_i\alpha$ (read “agent i is aware that α”). In this ways, modal operators allow us to construct formulas that describe the mind of an agent. Strictly speaking, in this subsection, where our goal is formally to define the language, these words do not yet possess any meaning, and we give the intended interpretation only as an aid to the reader. Their meaning will be governed by axioms, which are the subject of the next subsection.[5]

In addition to K_i and A_i, we will have a third modal operator denoted L_i in our language. Although we will not present axioms that govern the meaning of L_i until the next subsection, it is useful give an informal interpretation right now. We would like to use L_i to represent an

[5] If one augment propositional logic (instead of first order logic) with modal operators, what he gets will be propositional modal logic, which is the language used in most of the previous literature on unawareness. It should be clear that propositional modal logic is not rich enough to include formulas such as “I am not sure whether or not you are aware of something that I am not”.

alternative kind of knowledge that differs slightly from K_i. In particular, L_i stands for the kind of "know" that appears in the following English sentence: "If Madison had been aware of the right to universal suffrage, he would have known that it was important, and would have included it in the Bill of Rights". Here, "know" refers to knowledge in the benchmark case—a hypothetical case where Madison were not plagued by his unawareness of the right to universal suffrage. It is not the same as actual knowledge, because Madison *was* plagued by unawareness. In the previous literature, what we call "benchmark knowledge" (L_i) and "actual knowledge" (K_i) have been called "implicit knowledge" and "explicit knowledge", respectively. Although we do not think these names are ideal, they have become standard usage and we shall follow henceforth this convention.[6]

We now formally describe the language. Fix a set N of agents. Fix an infinite set X of *variables*, with typical elements $x, y, z, x_1, x_2, x_3, \ldots$, etc. Fix some vocabulary consisting of a set of relation symbols (*predicates*); e.g. $Tall(x)$ ("x is tall"), $Taller(x, y)$ ("x is taller than y"), etc. This generates a set Φ of *atomic formulas*, $P(x_1, \ldots, x_k)$, where P is a k-ary predicate and $x_1, \ldots, x_k \in X$ are variables. Our language $\mathcal{L}$ is the smallest set of *formulas* that satisfies the following conditions:[7]

- if $\phi \in \Phi$, then $\phi \in \mathcal{L}$;
- if $\alpha, \beta \in \mathcal{L}$, then $\neg\alpha \in \mathcal{L}$ and $\alpha \wedge \beta \in \mathcal{L}$;
- if $\alpha \in \mathcal{L}$ and $x \in X$, then $\forall x\alpha \in \mathcal{L}$;
- if $\alpha \in \mathcal{L}$ and $i \in N$, then $K_i\alpha \in \mathcal{L}$ and $A_i\alpha \in \mathcal{L}$ and $L_i\alpha \in \mathcal{L}$.

We use the following standard abbreviations:

- $\alpha \vee \beta$ for $\neg(\neg\alpha \wedge \neg\beta)$;
- $\alpha \rightarrow \beta$ for $\neg\alpha \vee \beta$;
- $\alpha \leftrightarrow \beta$ for $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$;
- $\exists x\alpha$ for $\neg\forall x\neg\alpha$;
- $U_i\alpha$ for $\neg A_i\alpha$.

Finally, we require that there is a special unary predicate called E. The intended interpretation of $E(x)$ is "x is real". The meaning of "real" will depend on the specific application. For example, a founding father writing the Bill of Rights may come up with the following

[6] We believe the names "implicit knowledge" and "explicit knowledge" obscure the fact that L_i is merely an conceptual tool, and is used only as an intermediate step to define K_i, which in turn is our ultimate interest.

[7] To be precise, we should have included "(" and ")" as building blocks of our language; we chose not to do so, and use parentheses only in cases of likely confusion.

rights: freedom of speech, freedom to bear arms, freedom of choosing one's own fate.... Upon further reflection, he may realize that only the first two are "real" rights, while the third one is merely an artificially-created concept. For an agent trying to enumerate animals, on the other hand, horses and cows are "real", while unicorns are not.

2.2 The Axiom System

Before we state our axioms, we need one more definition. We say that a variable x is *free* in the formula α if it does not fall under the scope of a quantifier $\forall x$.[8] For example, x is free in $\forall y\, Taller(x,y)$ but not in $\forall x \forall y\, Taller(x,y)$.

We are now ready to present our axiom system, called *AWARE*, for the language $\mathcal{L}$. As we explained earlier, an axiom system contains both axioms and inference rules. We shall group the axioms and inference rules into several different categories. The first category is borrowed directly from propositional logic, and is common to all axiom systems in this literature:

PC all propositional tautologies are axioms[9]
MP from α and $\alpha \rightarrow \beta$ infer β

PC (propositional calculus) is a set of axioms: propositional tautologies include formulas such as $\alpha \vee \neg\alpha$ and $\big(Tall(x) \wedge \exists x\,(x,y)\big) \rightarrow \exists x\, Taller(x,y)$; **MP** (*modus ponens*) is an inference rule, and says that if α and $\alpha \rightarrow \beta$ are provable formulas, then so is β. Note that any formulas may be substituted for α and β, and any variables for x and y.

The second category governs the universal quantifier $\forall$:

E1 for any variable x, $\forall x\, E(x)$ is an axiom
E2 for any formula α and variables x and y, $\forall x\alpha \rightarrow \big(E(y) \rightarrow \alpha[y/x]\big)$ is an axiom
E3 for any formulas α and β and variable x, $\forall x(\alpha \rightarrow \beta) \rightarrow (\forall x\alpha \rightarrow \forall x\beta)$ is an axiom
E4 for any formula α and variable x that is not free in α, $\alpha \leftrightarrow \forall x\alpha$ is an axiom
UG from α infer $\forall x\alpha$

[8] More formally, we define inductively what it is for a variable to be free in α:

- if ϕ is an atomic formula $P\,(x_1, \ldots, x_k)$, then each x is free in a formula;
- x is free in $\neg\alpha$, $K_i\alpha$, $A_i\alpha$, and $L_i\alpha$ iff x is free in α;
- x is free in $\alpha \wedge \beta$ iff x is free in α or β;
- x is free in $\forall y\alpha$ iff x is free in α and x is different from y.

Axiom **E1** can be rewritten as $\neg\exists x \neg E(x)$, which gives an interpretation to the existential quantifier in terms of the predicate E. Since there are at least two different ways to interpret the existential quantifier, **E1** is an important axiom in the sense that it clarifies which of the two interpretations is adopted in our system. Consider the following sentence:

> "There exist rights that have not been included in the Bill of Rights—think about the freedom to choose one's own fate."

Depending on how we interpret the word "exist", one may or may not agree that "the freedom to choose one's own fate" is an appropriate example in the above sentence. In particular, if we interpret the word "exist" according to **E1**, then we would likely regard that example as inappropriate, because that freedom is not a real right at all. However, one can conceive another interpretation of the word "exist" that would make that example appropriate. These two interpretations correspond to what logicians call *actualist existence* and *possibilitist existence*, respectively. We consider the possibilitist interpretation as less useful in economics. The reason, roughly speaking, is that most possibilitist axiom systems we are aware of lead to constant-domain structures, which is an especially restrictive property for economic models. We shall return to this point in Section 6, when we compare our current paper with Halpern and Rego (2006) and Sillari (2006).

In axiom **E2**, $\alpha[y/x]$ is the same formula as α with free y replacing every free x.[10] To understand axiom **E2**, one can consider the following conversation:

> FATHER: "One difference between horses and goats is that horses do not have horns."
>
> SON: "But unicorns have horns."
>
> FATHER: "Unicorns exist only in fantasy stories. What I meant was: *real* horses do not have horns."

The "E(y)" part of **E2** captures the father's qualification in his second statement, where α is "if x is a horse than x does not have horns". The basic idea is that the quantifier $\forall$ ranges only over "real" things. **E3** is straightforward. In **E4**, if x is not free in α, then adding $\forall x$ at the beginning of α does not change the meaning; for example, "for all things, Aumann is an economist" has the same information content as "Aumann is an economist" (although the former is rather awkward English). To understand **UG**, consider a formula such as "x is either tall or not tall". Suppose we have managed to find a proof for it by means of other axioms and inference rules. We would like to make sure that the formula "for all x, x is either tall or not tall" is also provable; and **UG** is an inference rule that will help make sure this.

The third category governs the meaning of explicit knowledge:

K for any formula α, $K_i\alpha \leftrightarrow (A_i\alpha \wedge L_i\alpha)$ is an axiom.

[10] So for example $\big(E(y) \wedge \forall x \exists z P(x,y)\big) \to \exists z P(y,y)$ is an axiom, but $\big(E(y) \wedge \forall x \exists y P(x,y)\big) \to \exists y P(y,y)$ is not.

We have already alluded to the idea behind **K** in Section 2.1: an agent explicitly knows a fact if and only if he implicitly knows it and he is not plagued by unawareness problems.

The fourth category governs the meaning of awareness:

A1 for any formula α that contains no free variables, $A_i\alpha$ is an axiom;
A2 for any formulas α and β, $(A_i\alpha \wedge A_i\beta) \rightarrow A_i(\alpha \wedge \beta)$ is an axiom;
A3 for any formulas α and β, if every variable x that is free in β is also free in α, then $A_i\alpha \rightarrow A_i\beta$ is an axiom.

To see that these three axioms capture the idea that awareness is object-based (alluded to in Section 1.2), one may heuristically think of a free variable as referring to some specific object, while a variable that is bound by a $\forall$ quantifier refers to generic objects. With this heuristic understanding, our idea that unawareness of a fact must arise from unawareness of specific objects referred to in the fact will have three implications, each correspond to one of the three axioms: if a fact does not refer to any specific objects, an agent will be aware of it (**A1**); if the agent is aware of two facts, then he is aware of a more complicated fact which is the conjunction of the two (**A2**); and if an agent is aware of a fact that refers to a collection of specific objects, then he is also aware of a fact that refers to only a subcollection of them (**A3**). These three implications, combined together, also characterize the idea that unawareness of a fact must arise from unawareness of specific objects referred to in the fact.

The last category governs the meaning of implicit knowledge:

L $L_i(\alpha \rightarrow \beta) \rightarrow (L_i\alpha \rightarrow L_i\beta)$;
LN from α infer $L_i\alpha$;
UGL from $\alpha_1 \rightarrow L_i\big(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\beta)\cdots\big)$, where $h \geq 0$, infer $\alpha_1 \rightarrow L_i\big(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\forall x\beta)\cdots\big)$, provided that x is not free in $\alpha_1, \ldots, \alpha_h$.

These axioms suggest that our agents are very powerful reasoners indeed, at least implicitly. Both **L** and **LN**, with explicit knowledge replacing implicit knowledge, are present in (the axiom system that underlies) all standard state-space models. In this sense they are standard assumptions in economics. **L** says that the agent can apply the inference rule *modus ponens* in his head (at least when he is not plagued by unawareness problems). **LN** says that our agents implicitly know all provable formulas of *AWARE*, even formulas that no one has ever written down, let alone found a proof for and published in a journal.

UGL is a bit mouthful. To understand what it says, it may be useful to put $h = 0$ and simplify it to "from $L_i\beta$ infer $L_i\forall x\beta$," which in turn takes a form similar to **UG**.

From these axioms and inference rules we can define the set of provable formulas. A formula can qualify as provable directly because it is an axiom, or it can qualify indirectly by association. The process of showing that a formula qualifies as a provable formula is called a "proof." Formally, a *proof* is a finite sequence of formulas, each of which is either an axiom or follows from preceding formulas in the sequence by applying an inference rule.

A proof of α is such a sequence whose last formula is α. A formula is a provable formula iff it has a proof.

2.3 The Object-Based Unawareness Structures

We now present the class of structures $\mathcal{M}$ (together with a truth-value assignment rule); we shall show (Theroem 1) that this class of structures is axiomatized by the axiom system *AWARE* presented in Section 2.2 above.

An object-based unawareness (OBU) structure is a tuple $M = (\mathcal{W}, D, \{D_w\}, \mathcal{P}_1, \ldots, \mathcal{P}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$, where:

- $\mathcal{W}$ is a set of possible worlds, with typical element w;
- D is a set of objects;
- for each w, D_w is a non-empty subset of D, containing objects that are "real" in w;
- $\mathcal{P}_i$ is agent i's *possibility correspondence*; it is a mapping that assigns to each w a subset of $\mathcal{W}$, containing worlds that agent i considers possible when the true world is w;
- $\mathcal{A}$ is agent i's *awareness correspondence*; it is a mapping that assigns to each w a subset of D, containing objects that agent i is aware of in w;
- π is an *assignment* at each w of a k-ary relation $\pi(w)(P) \subseteq D^k$ to each k-ary predicate P.

Intuitively, the assignment π describes the properties of each object; these can differ across worlds, so that for example Alice could be taller than Bob in world w_1, while Bob is taller than Alice in world w_2. Let $\mathcal{M}$ be the class of all OBU structures.

The way we intend to use an OBU structure is very standard, and will be formally captured in the truth-value assignment rule. Before we present the rule, let's go through two simple examples first.

As the first example, suppose John is an element in D, and *tall* is one of the predicates. To determine whether or not, in world w, agent i knows that John is tall, we first construct the event that John is tall, which is $E := \{w | John \in \pi(w)(tall)\}$. We then ask two questions: in world w, (1) does i implicitly know that John is tall ($\mathcal{P}_i(w) \subset E$)? and (2) is i aware of John ($John \in \mathcal{A}_i(w)$)? If both answers are affirmative, then i knows that John is tall in world w.

As another example, suppose we want to determine whether or not, in world w, i knows that everyone is tall. Once again, we first construct the event that everyone is tall, which is $E := \{w | D_w \subset \pi(w)(tall)\}$. Note that we only count those people who are "real"—for example, in a world where Jesus has no son, we do not count "Jesus' son" even if he is an element in D. We then ask: in world w, does i implicitly know that everyone is tall? If the answer is affirmative, then i knows that everyone is tall in world w. Note that we do not need to ask the awareness question, because no specific person is referred to in the fact "everyone is tall", and hence by assumption there will be no unawareness problem.

These will all be formally captured by our truth-value assignment rule. Let a *valuation* V on an OBU structure M be a function that assigns a member of D to each variable x. Intuitively, $V(x)$ describes the object referred to by variable x, provided that it appears free in a given formula, just like how a name is associated to an object. The truth value of a formula depends on the valuation, just like whether or not "Bob is tall" depends on which person bears the name "Bob".

We say that the fact represented by the atomic formula $E(x)$ is true at state w of structure M under valuation V, and write

$$(M, w, V) \models E(x),$$

iff $V(x)$ is one of the objects in D_w. For facts represented by more complicated formulas, we use the following rules inductively:

$(M, w, V) \models P(x_1, \ldots, x_k)$ iff $\big(V(x_1), \ldots, V(x_k)\big) \in \pi(w)(P)$;

$(M, w, V) \models \neg\alpha$ iff $(M, w, V) \not\models \alpha$;

$(M, w, V) \models \alpha \wedge \beta$ iff $(M, w, V) \models \alpha$ and $(M, w, V) \models \beta$;

$(M, w, V) \models \forall x\alpha$ iff $(M, w, V') \models \alpha$ for every x-alternative V' of V such that $V'(x) \in D_w$;[11]

$(M, w, V) \models A_i\alpha$ iff $V(x) \in \mathcal{A}_i(w)$ for every x that is free in α;

$(M, w, V) \models L_i\alpha$ iff $(M, w', V) \models \alpha$ for all $w' \in \mathcal{P}_i(w)$;

$(M, w, V) \models K_i\alpha$ iff $(\mathcal{M}, w, V) \models A_i\alpha$ and $(\mathcal{M}, w, V) \models L_i\alpha$.

If α is true at every w in M under V, we say that α is valid in M under V, and write

$$(M, V) \models \alpha.$$

If α is valid in M under every V, we say that α is valid in M, and write

$$M \models \alpha.$$

If α is valid in every $M \in \mathcal{C} \subseteq \mathcal{M}$, we say that α is valid in $\mathcal{C}$, and write

$$\mathcal{C} \models \alpha.$$

[11] We say that V' is an x-alternative of V if, for every variable y except possibly x, $V'(y) = V(y)$.

2.4 The Characterization Theorem

Theorem 1 *The set of formulas $\alpha \in \mathcal{L}$ that are valid in $\mathcal{M}$ is exactly the set of provable formulas in AWARE.*

In logicians' terminology, *AWARE* is a sound and complete axiomatization of $\mathcal{M}$ in $\mathcal{L}$. The proof of Theorem 1 uses standard methodology.[12] We present the complete proof in the appendix.

3 An Application

Chung and Fortnow (2007) use a dynamic game with two players (a legislator who is to write the Bill of Rights, and a judge who is to interpret it 200 years later) to formalize the argument of those American founding fathers who opposed the inclusion of the Bill of Rights into the American Constitution. They prove that, in some parameter range, there is a unique equilibrium where the legislator, who *is not sure whether or not there are still other rights that he is unaware of*, optimally chooses to *not* to write the Bill of Rights. That is, he optimally chooses *not* to enumerate even those rights that he is aware of. The reason is that, in equilibrium, how the judge treats those rights not in the Bill depends on how elaborated the Bill is. The more elaborate the Bill is, the more likely that judge will rule that it is constitutional for the government to infringe rights that have not been listed.

They also prove that, even if the legislator adds the sentence

> "Any other rights not listed in this Bill are equally sacred and the government should not infringe them."

to the Bill, the equilibrium outcome will be the same.

Instead of reproducing the analysis of Chung and Fortnow (2007) here, let's focus on how one can use an OBU structure to model that legislator.

Consider the following object-based unawareness structure: There are two worlds, w_1 and w_2, and two rights, s and f, where s stands for "freedom of speech" and f stands for "freedom to choose one's own fate". The true state is w_2, where only s is "real". However, both s and f are "real" in the other world, w_1. Formally, it means $D_{w_1} = \{s, f\}$, and $D_{w_2} = \{s\}$. Suppose agent i is aware of only s in both worlds (i.e., $\mathcal{A}_i(w) = \{s\}$ for $w = w_1, w_2$). Then, in world w_1, and only in world w_1, there exists some object that the agent is unaware of. If we use P to stand for some arbitary property that both objects satisfy in both worlds (i.e., $\pi(w)(P) = \{s, f\}$ for $w = w_1, w_2$), then we have:

$$(M, w) \models \exists x U_i P(x),$$
$$(M, w) \models \neg \exists x U_i P(x).$$

[12] See, for example, the proof of Theorem 16.2 in Hughes and Cresswell (1996).

Suppose, in the true state w_2, the agent cannot distinguish w_1 and w_2 (i.e., $\mathcal{P}(w_2) = \{w_1, w_2\}$). Then he is not sure whether or not there exists some right that he is unaware of. I.e., he does not know for sure that there is no such a right:

$$(M, w_2) \models \neg K_i \neg \exists x U_i P(x);$$

and he does not know for sure that there is such a right:

$$(M, w_2) \models \neg K_i \exists x U_i P(x).$$

This lack of explicit knowledge is not due to unawareness, for he can comprehend both facts:

$$\begin{aligned} &(M, w_2) \models A_i\big(\neg \exists x U_i P(x)\big), \quad \text{and} \\ &(M, w_2) \models A_i\big(\exists x U_i P(x)\big). \end{aligned}$$

His lack of explicit knowledge is due to his lack of implicit knowledge—he does not (implicitly) know for sure the exact number of rights that really exist.

Note that this is an example with non-constant domains; i.e., D_w varies across different worlds. Non-constant domains are important in modelling agents who are not sure whether or not there exist things that they are unaware of. Consider the above example again, but suppose D_w is constant across different worlds. For example, suppose $D_w = \{s\}$ in both worlds. Then $K_i \neg \exists x U_i P(x)$ would have been true in both worlds. Alternatively, suppose $D_w = \{s, f\}$ in both worlds. Then $K_i \exists x U_i P(x)$ would have been true in both worlds. The possibility of non-constant domains in our structures arises from our adoption of the actualist existence axioms. In Halpern and Rego (2006) and Sillari (2006), where they adopt the possibilitist existence axioms, their structures are necessarily characterized by constant domains instead.

4 Dekel, Lipman, and Rustichini's Axioms

A first impression of some readers of this paper is that object-based unawareness structures violate DLR's AU Introspection Axiom. AU Introspection is represented by formulas of the form $U_i\alpha \to U_iU_i\alpha$. Indeed, every such formula is a provable formula of *AWARE* (and hence, by Theorem 1, it is valid in $\mathcal{M}$). To see why, first note that $U_i\alpha$ and α have the same free variables, so $A_iU_i\alpha \to A_i\alpha$ is a provable formula of *AWARE* (for all $\alpha \in \mathcal{L}$ and all i). By simple propositional reasoning, then, $U_i\alpha \to U_iU_i\alpha$ is also a provable formula of *AWARE*.

Consider however a similar formula: $U_i\alpha \to U_i\exists x U_i\alpha$. Suppose α stands for $H(x)$, "x is a human right". Then this formula reads "if an agent is unaware that free speech is a human right, then she is unaware that there is any human right that she is not aware of". Clearly we would not want this formula to be a provable formula in *AWARE* (or valid in $\mathcal{M}$): clearly we would like to be able to model agents who are unaware of some things, but aware (or even explicitly know) that there are things they are unaware of. To show that this formula is indeed not valid in $\mathcal{M}$ (and hence not a provable formula in *AWARE*), it suffices to show that its negation is true in some world w of some object-based unawareness structure $M \in \mathcal{M}$ under some valuation V. If α stands for $H(x)$, x is free in α but not in $\exists x U_i \alpha$. Then if $V(x) \notin \mathcal{A}_i(w)$, we have $(M, w, V) \models U_i \alpha$ but $(M, w, V) \not\models U_i \exists x U_i \alpha$, and so $(M, w, V) \models \neg(U_i \alpha \rightarrow U_i \exists x U_i \alpha)$.

Another DLR axiom is Plausibility, which is represented by formulas of the form

$$U_i \alpha \rightarrow (\neg K_i \alpha \wedge \neg K_i \neg K_i \alpha).$$

Again, every such formula is a provable formula in *AWARE*. This follows easily from **A3** and **K**.

DLR's third axiom is KU Introspection, which is represented by formulas of the form $\neg K_i U_i \alpha$. Such formulas are not provable formulas in *AWARE*. The basic reason is that there are no axioms in *AWARE* to preclude an agent knowing something that is actually false. (In this sense, instead of implicit and explicit knowledge, we should perhaps call L_i and K_i implicit and explicit *belief*.) So an agent may explicitly know/believe that she is unaware of something, even though she is actually aware of it. Adding the Truth Axiom (**T**: $L_i \alpha \rightarrow \alpha$), would make every instance of $\neg K_i U_i \alpha$ a provable formula.[13] In terms of our structures, **T** corresponds to the restriction that the possibility correspondences are reflexive: $w \in \mathcal{P}_i(w)$. To be more precise, let $\mathcal{M}^r$ be the class of object-based unawareness structures in which each $\mathcal{P}_i$ is reflexive; then the set of formulas that are valid in $\mathcal{M}^r$ is precisely the set of provable formulas of *AWARE*+**T**.

DLR's impossibility result is stated within the confines of standard state-space models, and they argue that Necessitation and Monotonicity are two characterizing features of those models. Both Necessitation and Monotonicity are restrictions imposed on their class of state-space models, and can be translated into restrictions on our OBU structures as well. Necessitation corresponds to the restriction that, for any given OBU structure M and valuation , if the formula α is true (i.e., $(M, w,) \models \alpha$) in every world w then the formula $K_i \alpha$ is also true (i.e., $(M, w,) \models K_i \alpha$) in every world w. Monotonicity corresponds to the restriction that, for any given OBU structure M and valuation , if the formula $\alpha \rightarrow \beta$ is true (i.e., $(M, w,) \models \alpha \rightarrow \beta$) in every world w then the formula $K_i \alpha \rightarrow K_i \beta$ is also true (i.e., $(M, w,) \models K_i \alpha \rightarrow K_i \beta$) in every world w. In general, our OBU structures do not satisfy these two restrictions. After all, our OBU structures are not standard state-space models.

5 Other Axioms of Interest

The axiom system *AWARE* (and, correspondingly, the class $\mathcal{M}$ of all OBU structures) can be thought of as imposing a minimal set of restrictions on the behavior of our language $\mathcal{L}$. Various additional assumptions have been imposed on models of knowledge and unawareness elsewhere in the literature. In this section, we shall discuss several such assumptions. In each

[13] Since our language has two knowledge operators, there are two ways to write the Truth Axiom. The stronger version is is $L_i \alpha \rightarrow \alpha$, which we adopted in the text. An alternative, weaker version is $K_i \alpha \rightarrow \alpha$. Here, even adding the weaker version suffices to make every instance of $\neg K_i U_i \alpha$ a provable formula.

case, we offer an axiomatic representation, and explain how it corresponds to a particular subclass of $\mathcal{M}$.

To begin with, the following axioms are standard in the economics literature, and are implicit in the partitional model of knowledge used in the vast majority of economic applications:

T for any formula α, $L_i\alpha \rightarrow \alpha$ is an axiom;
PI for any formula α, $L_i\alpha \rightarrow L_iL_i\alpha$ is an axiom;
NI for any formula α, $\neg L_i\alpha \rightarrow L_i\neg L_i\alpha$ is an axiom.

We have already come across the Truth Axiom **T** in Section 4. **PI** is the Axiom of Positive Introspection, and **NI** is the Axiom of Negative Introspection. Note that all three are stated in terms of implicit knowledge L_i instead of explicit knowledge K_i. These axioms have been interpreted by some as rationality requirements on the agents, but generally they are considered to be unrealistically strong.

As before, we say that an agent's possibility correspondence $\mathcal{P}_i$ is reflexive if $w \in \mathcal{P}_i(w)$ for all w. We say that it is *transitive* if $x \in \mathcal{P}_i(w)$ and $y \in \mathcal{P}_i(x)$ imply $y \in \mathcal{P}_i(w)$ for all w, x, y; and *Euclidean* if $x \in \mathcal{P}_i(w)$ and $y \in \mathcal{P}_i(w)$ imply $y \in \mathcal{P}_i(x)$ for all w, x, y. Let $\mathcal{M}^r$, $\mathcal{M}^t$, and $\mathcal{M}^e$ denote the subclasses of $\mathcal{M}$ in which all $\mathcal{P}_i$'s are reflexive, transitive, and Euclidean, respectively. We shall also use, for example, $\mathcal{M}^{re}$ to denote the subclass of $\mathcal{M}$ in which all $\mathcal{P}_i$'s are reflexive *and* Euclidean. The following straightforward extension of Theorem 1 formalizes the notion that reflexivity corresponds to **T**, transitivity to **PI**, and Euclideaness to **NI**:

Theorem 2 *The set of formulas $\alpha \in \mathcal{L}$ that are valid in $\{M^r, \mathcal{M}^t, \mathcal{M}^e, \mathcal{M}^{rt}, \mathcal{M}^{re}, \mathcal{M}^{re}, \mathcal{M}^{rte}\}$ is exactly the set of provable formulas in AWARE+* ***T,PI,NI, TPI,TNI,PINI,TPINI****.*

One may also be interested in an axiom that says every agent knows what he is aware of:

KA for any formula α, $A_i\alpha \rightarrow K_iA_i\alpha$ is an axiom.

A related axiom, A-Introspection ($A_i\alpha \leftrightarrow K_iA_i\alpha$), appears in Heifetz, Meier, and Schipper (2006). Note that **KA** and **A3** imply A-Introspection. *AWARE*+**KA** corresponds to the subclass of $\mathcal{M}$ in which the possibility correspondences satisfy the following restriction: for any w and any $w' \in \mathcal{P}_i(w)$, $\mathcal{A}_i(w) \subseteq \mathcal{A}_i(w')$.

In the presence of **A3** and **K**, it is straightforward to show that **KA** is equivalent to:

LA1 for any formula α, $A_i\alpha \rightarrow L_iA_i\alpha$ is an axiom.

Inspired by **LA1**, some may be tempted to add its "mirror image" as well:

LA2 for any formula α, $U_i\alpha \rightarrow L_iU_i\alpha$ is an axiom.

LA2 has appeared in some earlier studies on unawareness.[14] We cannot, however, think of any justification for it,[15] other than the purely aesthetic fact that it looks similar to **LA1**. It is straightforward to show that $AWARE$+**LA1**+**LA2** corresponds to the subclass of $\mathcal{M}$ where, for any w and any $w' \in \mathcal{P}(w)$, $\mathcal{A}(w) = \mathcal{A}(w')$.

6 Related Literature

In this section we shall discuss some of the important contributions to the literature on unawareness.[16] All of these papers share a common feature: unawareness is associated with events/facts instead of with objects/things.

In an early paper, Fagin and Halpern (1988) take as their starting point the language of propositional modal logic, and add unawareness modal operators to allow for U-sentences. To construct models that do not preclude U-sentences, they augment the standard Kripke structures[17] with an unawareness function. The unawareness function associates with each state a subset of formulas, listing the facts that the agent is unaware of in that state. They impose no restriction on the unawareness function, so the agent could be aware of a formula but unaware of its negation. They also consider an assumption that awareness is *closed under subformulas*, which rules out this possibility. They provide an axiomatization for their structures analogous to our Theorem 1.

Modica and Rustichini (1999) provide the first treatment of unawareness in the economics literature that avoids DLR's critique (and also address concerns raised in an earlier paper of their own, Modica and Rustichini (1994)). Their models, called *generalized standard models*, distinguish between an objective set of possible worlds and a subjective subset, with the latter used to represent facts that the agent is aware of. Halpern (2001) shows that generalized standard models can be viewed as special cases of those in Fagin and Halpern (1988), with appropriate restrictions on the awareness function. Li (2006) uses a similar technique to model multi-agent unawareness; it should be noted that the extension to multiple agents is not trivial in this context.

In a strikingly original paper, Heifetz, Meier, and Schipper (2006) deal with the extension to the multi-agent case in a different way. They work with a partially ordered *set of sets* of possible worlds, where the ordering represents the expressive power of each set. For instance, if there are only two primitive propositions of interest, their structure consists of four sets of sets, with the most expressive one describing situations involving both propositions, two less expressive sets describing situations involving the first and the second propositions respectively, and the least expressive set describing only situations that involve neither. These sets are used to represent the awareness of agents. In a companion paper, Heifetz, Meier,

[14] It appears, for example, as Axiom **A12** in Halpern (2001).

[15] Or perhaps we should say we are not *aware* of any justification for it.

[16] A comprehensive bibliography can be found on Burkhard Schipper's website: `http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm`

[17] A *Kripke structure* is, roughly, a more general version of the partitional information structure, together with a function that specifies which primitive propositions hold in which states.

and Schipper (2007) provide an axiomatization for their structures.

None of the papers discussed so far allows us to model agents' reasoning about their possible lack of awareness. Two recent papers work with languages that include EU-sentences. Halpern and Rego (2006) use second-order modal logic, augmenting the language of Fagin and Halpern (1988) by including quantifiers *over formulas*. The resulting language includes formulas such as $\forall x K_i A_j x$, to be read as "agent i knows that agent j is aware of every formula". More closely related to our current paper, Sillari (2006) uses a language that is identical to ours.

One difference between these two papers and ours is that they have very different axioms for the existential quantifier. In particular, their axioms correspond to what logicians call the *possibilitist* interpretation of existence, whereas ours correspond to the *actualist* interpretation. Although both kinds of axioms have their proponents, we believe possibilitist existence is less useful when it comes to constructing economic models. The reason, roughly speaking, is that most possibilitist axiom systems we are aware of come with the *Barcan Formula*[18], which, when coupled with other axioms familiar to economists, will have undesirable implications. To illustrate this, let's consider what would have happened had we adopted the possibilitist axioms as well. By this, we mean replacing our Axiom **E1** with the Barcan Formula:

BF for any formula α and variables x, $\forall x L_i \alpha \rightarrow L_i \forall x \alpha$ is an axiom,

and replacing our Axiom **E2** with:

E2' for any formula α and variables x and y, $\forall x \alpha \rightarrow \alpha[y/x]$ is an axiom.

The class of structures that is axiomatized by this new axiom system is exactly those object-based unawareness structures where $D_w = D$ in every world w. If we further add Axioms **LA1** and **LA2**—which, as we explained in Section 5, appeal to many economists—then the resulting subclass of $\mathcal{M}$ will have a very undesirable feature. In any structure within this subclass, an agent either knows for sure that there exists something he is unaware of, or knows for sure that there is nothing he is unaware of—but he can never be uncertain. If he were to assign a probability to the event that there exists something he is unaware of, then that probability would have to be either 0 or 1—it could not lie strictly between.

As Sillari (2006) points out, this very problem arises in Halpern and Rego (2006): "The second-order logic of Halpern and Rego also requires the Barcan to be validated, hence does not lend itself to model knowledge as high-probability operators." That is, once Halpern and Rego incorporate axioms analogous to **LA1** and **LA2** into their axiom system, sentences such as "the agent is not sure whether or not there are still things that he is unaware of" or "I am not sure whether or not you are aware of something that I am not" will become contradictory in their resulting structures—they must be false in every world of every structure.

[18] It is named after the philosopher and logician Ruth Barcan Marcus, the founder of first-order modal logic.

Although Sillari (2006) also adopts the possibilitist axioms for the existential quantifier, his axiom system is an exception in that it does not contain the the Barcan formula, as he has very different axioms for implicit knowledge. His weaker axioms on implicit knowledge lead to a class of structures very different from our OBU structures. Roughly speaking, our OBU structures are generalizations of Kripke structures, which are more familiar to economists; while his structures are generalizations of the neighborhood semantics.

Another, more important, difference between Halpern and Rego (2006) and Sillari (2006) and our current paper lies in the way unawareness is modelled. Both of them use the same approach as Fagin and Halpern (1988) by introducing an unawareness function that assigns to each agent at each possible world a list of those formulas that agent is unaware of—we call this the *semi-syntactic approach*. In our object-based approach, on the other hand, we provide a foundation for awareness of formulas in terms of awareness of objects.

We believe the object-based approach offers an advantage over the semi-syntactic approach. Logicians like to preserve a clear distinction between the extra-linguistic reality (which we can think of in our structure as $\mathcal{W}$, D, D_w, $\mathcal{P}_i$, and $\mathcal{A}_i$) and the *semantics* (the truth-value assignment rule) which maps the language into this reality. This distinction is cut by the semi-syntactic approach, which explicitly uses the language to represent part of the reality (specifically, the awareness of the agents). Why does this matter? In the semi-syntactic approach, any restrictions that are imposed on the awareness function (Halpern and Rego (2006) consider several) must of course be expressed linguistically, and correspond closely to equivalent axioms in the axiom system. But in the object-based approach, the (non-linguistic) awareness function and the assumptions we make about it look very different from the corresponding axioms governing the behavior of the awareness operator: this gives us two different viewpoints from which to assess the reasonableness of our underlying model of awareness.

At the risk of setting up a strawman, consider as an analogy two different ways of modelling knowledge: first, the standard approach, where we have a set of possible worlds, and a possibility correspondence for each agent describing, at each world, which worlds the agent consider possible; second, a semi-syntactic approach, where instead of the possibility correspondence we have a knowledge function which simply lists the set of formulas each agent knows at each world. Just as various assumptions about the possibility correspondence (that it is reflexive, transitive, etc.) correspond to various axiomatizations of the properties of knowledge (in some appropriate language), restrictions could be imposed on the knowledge function to derive similar equivalence results. But it is clear that the standard approach offers us two distinct perspectives on the concept of knowledge, and potentially a better understanding of it, while the semi-syntactic approach offers only one.

Finally, we should also mention the contribution of Feinberg (2004), who adopts an ingenious meta-approach to the problem: instead of attempting to express unawareness directly within the formal language, he describes unawareness implicitly by describing which subsets of the language make up each agent's subjective world view.

7 Conclusion

The goal of this paper is to construct a user-friendly model that allows us to express EU-sentences such as “the agent is not sure whether or not there are still things that he is unaware of”. Instead of trying to assign truth values to these EU-sentences within existing unawareness models in the literature, and worrying about whether or not the truth-value-assignment rule is consistent with some set of “reasonable” axioms, we started with an explicitly list of axioms, and then constructed the class of structures (together with a truth-value-assignment rule) that is axiomatized by exactly those axioms. As an application, we explained how our structures can be use to model those American founding fathers who were opposed to the inclusion of the Bill of Rights into the constitution.

Appendix A:

In this appendix we shall prove Theorem 1. Throughout this proof, **PC** and **MP** will be used too often for us to acknowledge everytime. Hence we shall often refrain from citing their names when we use them.

As usual in this literature, the proof involves two steps: the soundness part and the completeness part. The soundness part says that all provable formulas of *AWARE* are valid in $\mathcal{M}$. The completeness part says the converse is also true.

Lemma 1 *Every provable formula $\alpha \in \mathcal{L}$ of AWARE is valid in $\mathcal{M}$.*

Proof: We shall prove that each axiom is valid in every $M \in \mathcal{M}$, at every w, and under every V; and that each inference rule preserves validity. We shall, however, skip the parts of **PC** and **MP**.

For **E1**, notice that for every x-alternate V' of V such that $V'(x) \in D_w$, we have $(M, w, V') \models E(x)$, which implies $(M, w, V) \models \forall x E(x)$.

For **E2**, suppose $(M, w, V) \models \forall x\alpha$ and $(M, w, V) \models E(y)$ but $(M, w, V) \not\models \alpha[y/x]$. Let V' be the x-alternative of V such that $V'(x) = V(y)$. Then we have both $(M, w, V') \not\models \alpha$ and $V'(x) \in D_w$, contradicting that $(M, wV) \models \forall x\alpha$.

For **E3**, suppose $(M, w, V) \models \forall x(\alpha \to \beta)$ and $(M, w, V) \models \forall x\alpha$. Then, for any x-alternative V' of V such that $V'(x) \in D_w$, we have both $(M, w, V') \models \alpha \to \beta$ and $(M, w, V') \models \alpha$, which implies $(M, w, V') \models \beta$, which in turn implies $(M, w, V) \models \forall x\beta$.

For **E4**, notice that if x is not free in α, then $(M, w, V) \models \alpha$ iff $(M, w, V') \models \alpha$ for any x-alternative V' of V.

For **UG**, suppose $(M, w, V) \not\models \forall x\alpha$. Then there exists some x-alternative V' of V such that $V'(x) \in D_w$ and $(M, w, V') \not\models \alpha$, which implies that the formula α is not valid in $\mathcal{M}$.

For **K**, it follows directly from the truth-value-assignment rule in Section 2.3.

For **A1**, notice that if α contains no free variables, then $V(x) \in \mathcal{A}(w)$ for every x free in α, and hence we have $(M, w, V) \models A_i\alpha$.

For **A2**, suppose $(M, w, V) \models A_i\alpha$ and $(M, w, V) \models A_i\beta$. Then $V(x) \in \mathcal{A}(w)$ for every free x in α and every free x in β, and hence also for every free x in $\alpha \wedge \beta$, and hence we have $(M, w, V) \models A_i(\alpha \wedge \beta)$.

For **A3**, suppose $(M, w, V) \models A_i\alpha$. Then $V(x) \in \mathcal{A}(w)$ for every free x in α, and hence also for every free x in β, and hence we have $(M, w, V) \models A_i\beta$.

For **L**, suppose $(M, w, V) \models L_i(\alpha \to \beta)$ and $(M, w, V) \models L_i\alpha$ but $(M, w, V) \not\models L_i\beta$. Then there exists some $w' \in \mathcal{P}_i(w)$ such that $(M, w', V) \models \alpha \to \beta$ and $(M, w', V) \models \alpha$ but $(M, w', V) \not\models \beta$, contradicting the truth-value-assignment rule in Section 2.3.

For **LN**, suppose $(M, w, V) \not\models L_i\alpha$. Then there exists some $w' \in \mathcal{P}_i(w)$ such that $(\mathcal{M}, w', V) \not\models \alpha$, which implies the formula α is not valid in $\mathcal{M}$.

For **UGL**, suppose $(M, w, V) \not\models \alpha_1 \to L_i\big(\alpha_2 \to \cdots \to L_i(\alpha_h \to L_i\forall x\beta)\cdots\big)$. Then there is a sequence $w_1, \ldots, w_{h+1}$ such that $w_1 = w$, $(M, w_k, V) \models \alpha_k$ for $1 \leq k \leq h$, and $(M, w_{h+1}, V) \not\models \forall x\beta$. Moreover, there eixsts some x-alternative V' of V such that $V'(x) \in D_{w_{h+1}}$ and $(M, w_{h+1}, V') \not\models \beta$. Since x is not free in each α_k, we have $(M, w_k, V') \models \alpha_k$ for $1 \leq k \leq h$, which implies $(M, w, V') \not\models \alpha_1 \to L_i\big(\alpha_2 \to \cdots \to L_i(\alpha_h \to L_i\beta)\cdots\big)$, which in turn implies the formula $\alpha_1 \to L_i\big(\alpha_2 \to \cdots \to L_i(\alpha_h \to L_i\beta)\cdots\big)$ is not valid in $\mathcal{M}$. ■

By the truth-value-assignment rule in Section 2.3, the formula $\alpha \wedge \neg\alpha$ is not valid in $\mathcal{M}$, and hence by Lemma 1 is not a provable formula in *AWARE*. That there are some formulas that are not provable in *AWARE* means that the system is "consistent" in logicians' terminology. More importantly, it implies that it cannot be the case that both α and $\neg\alpha$ are provable formulas of *AWARE*. This observation will be used in subsequent proofs.

As usual, the proof of the completeness part involves the construction of a structure $M \in \mathcal{M}$, called the canonical structure, and a valuation V, such that every formula $\alpha \in \mathcal{L}$ that is valid in M under V is a provable formula of *AWARE*. Completeness then follows from the fact that any formula $\alpha \in \mathcal{L}$ that is valid in $\mathcal{M}$ must also be valid in M under V.

We say that a formula $\alpha \in \mathcal{L}$ is *AWARE-consistent* if $\neg\alpha$ is not a provable formula of *AWARE*. We say that a finite list of formulas $\{\alpha_1, \ldots, \alpha_k\} \subset \mathcal{L}$ is *AWARE*-consistent if the formula $\alpha_1 \wedge \ldots \wedge \alpha_k$ is *AWARE*-consistent. We say that an infinite list of formulas is *AWARE*-consistent if every finite sublist of it is *AWARE*-consistent.

We say that a list of formulas is *maximal* if, for every formula $\alpha \in \mathcal{L}$, either α or $\neg\alpha$ is in the list. We say that a list of formulas is *maximal AWARE-consistent* if it is both maximal and *AWARE*-consistent.

It is a standard result that if α is a provable formula of *AWARE*, then it is in every maximal *AWARE*-consistent list.

We say that a list Γ of formulas possesses the $L\forall$-property if

1. for every formula α and variable x, there is some variables y such that the formula $E(y) \wedge (\alpha[y/x] \to \forall x\alpha)$ is in Γ;
2. for any formulas $\beta_1, \ldots, \beta_h$ ($h \geq 0$) and α, and every variable x that is not free in $\beta_1, \ldots, \beta_h$, there is some variable y such that the formula $L_i\Bigg(\beta_1 \to \cdots \to L_i\Bigg(\beta_h \to L_i\big(E(y) \to \alpha[y/x]\big)\Bigg)\cdots\Bigg) \to L_i\big(\beta_1 \to \cdots \to L_i(\beta_h \to L_i\forall x\alpha)\cdots\big)$ is in Γ.

Lemma 2 *If formula α is AWARE-consistent, then there is an AWARE-consistent list Γ of formulas with the $L\forall$-property such that $\alpha \in \Gamma$.*

To prove Lemma 2, we need another lemma first.

Lemma 3 *The formula $\exists y(\theta[y/x] \to \forall x\theta)$ is a provable formula of AWARE.*

Proof: By **E2**, the formula $\big(E(x) \wedge \forall y\theta[y/x]\big) \rightarrow (\theta[y/x])[x/y]$ is a provable formula. Notice that $(\theta[y/x])[x/y]$ gives us back θ. Therefore, by **UG** and **E3**, the formula $\forall x E(x) \rightarrow \forall x(\forall y\theta[y/x] \rightarrow \theta)$ is a provable formula. By **E1** and **E3**, the formula $\forall x\forall y\theta[y/x] \rightarrow \forall x\theta$ is a provable formula. But x is not free in $\forall y\theta[y/x]$ anymore, and hence by **E4**, the formula

$$\forall y\theta[y/x] \rightarrow \forall x\theta \tag{1}$$

is a provable formula.

Given (1), it suffices to prove that the formula

$$\forall y\neg(\theta[y/x] \rightarrow \forall x\theta) \rightarrow \neg(\forall y\theta[y/x] \rightarrow \forall x\theta) \tag{2}$$

is a provable formula.

By **PC**, both formulas $\neg(\theta[y/x] \rightarrow \forall x\theta) \rightarrow \theta[y/x]$ and $\neg(\theta[y/x] \rightarrow \forall x\theta) \rightarrow \neg\forall x\theta$ are provable formulas. By **UG** and **E3**, both formulas $\forall y\neg(\theta[y/x] \rightarrow \forall x\theta) \rightarrow \forall y\theta[y/x]$ and $\forall y\neg(\theta[y/x] \rightarrow \forall x\theta) \rightarrow \forall y\neg\forall x\theta$ are also provable formulas. Since y is not free in $\neg\forall x\theta$, by **E4**, we have (2) as needed. ■

Proof of Lemma 2: Assume that all variable x are enumerated, and similarly for all formulas of the form $\forall x\theta$, and similarly for all formulas of the form $L_i\big(\xi_i \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)$ with $h \geq 0$ and x not free in $\xi_1, \ldots, \xi_h$.

Define a sequence of lists of formulas $\Gamma_0, \Gamma_1, \ldots$ as follows: $\Gamma_0 = \{\alpha\}$. Given Γ_n, we define Γ_{n+1} in two steps.

Step 1: We first extend Γ_n to Γ_n^+. Let $\forall x\theta$ be the $n+1$st formula of this form. Let y be the first variable that does not appear in Γ_n and θ, and define

$$\Gamma_n^+ = \Gamma_n \cup \{E(y), \theta[y/x] \rightarrow \forall x\theta\}.$$

We claim that, as long as Γ_n is *AWARE*-consistent, Γ_n^+ is *AWARE*-consistent as well. Suppose not. Then the formula $\beta \rightarrow \big(E(y) \rightarrow \neg(\theta[y/x] \rightarrow \forall x\theta)\big)$, where β denote the (finite) conjunction of all formulas in Γ_n, is a provable formula. By **UG**, **E3**, and **E4** (applicable because y does not occur in β), the formula $\beta \rightarrow \big(\forall y E(y) \rightarrow \forall y\neg(\theta[y/x] \rightarrow \forall x\theta)\big)$ is a provable formula. By **E1**, the formula $\beta \rightarrow \forall y\neg(\theta[y/x] \rightarrow \forall x\theta)$ is a provable formula. By Lemma 3, the formula $\neg\beta$ is a "theroem," contraditing the presumption that Γ_n is *AWARE*-consistent.

Step 2: We next extend Γ_n^+ to Γ_{n+1}. Let $L_i\big(\xi_i \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)$ be the $n+1$st formula of this form. Let y be the first variable that does not appear in Γ_n^+ and $\xi_1, \ldots, \xi_h$, and define $\Gamma_{n+1} = \Gamma_n^+ \cup \{L_i\Bigg(\xi_1 \rightarrow \cdots \rightarrow L_i\Big(\xi_h \rightarrow L_i\big(E(y) \rightarrow \theta[y/x]\big)\Big)\cdots\Bigg) \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)\}$.

We claim that, as long as Γ_n^+ is *AWARE*-consistent, Γ_{n+1} is *AWARE*-consistent as well.

Suppose not. Then both formulas

$$\beta \to L_i\Big(\xi_1 \to \cdots \to L_i\big(\xi_h \to L_i(E(y) \to \theta[y/x])\big) \cdots\Big) \tag{3}$$

and

$$\beta \to \neg L_i(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta) \cdots), \tag{4}$$

where β denote the (finite) conjunction of all formulas in Γ_n^+, are provable formulas.

Since y does not appear in Γ_n^+, by **UGL** (putting $n = h + 1$), from (3) we infer that the formula

$$\beta \to L_i\Big(\xi_1 \to \cdots \to L_i\big(\xi_h \to L_i\forall y(E(y) \to \theta[y/x])\big) \cdots\Big) \tag{5}$$

is a provable formula.

By **E3**, **LN**, and **L**, the formula $L_i\forall y(E(y) \to \theta[y/x]) \to (L_i\forall y E(y) \to L_i\forall y\theta[y/x])$ is a provable formula. Since by **E1** and **LN**, the formula $L_i\forall y E(y)$ is also a provable formula, we infer that the formula $L_i\forall y(E(y) \to \theta[y/x]) \to L_i\forall y\theta[y/x]$ is a provable formula. By using **LN** and **L** repeatedly (for h times to be exact), we infer that the formula $L_i\Big(\xi_1 \to \cdots \to L_i\big(\xi_h \to L_i\forall y(E(y) \to \theta[y/x])\big) \cdots\Big) \to L_i(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall y\theta[y/x]) \cdots)$ is a provable formula. From this, together with (5), we infer that $\beta \to L_i(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall y\theta[y/x]) \cdots)$ is a provable formula. From this, together with (4), we infer that $\neg\beta$ is a provable formula, contradicting the presumption that Γ_n^+ is *AWARE*-consistent.

We can now let Γ be the union of all Γ_n's. Since Γ_0 is *AWARE*-consistent, Γ is also *AWARE*-consistent. And Γ will have the $L\forall$-property by construction. ■

Lemma 4 *If an AWARE-consistent list Γ of formulas possess the $L\forall$-property, then there is a maximal AWARE-consistent list Δ fo formulas with the $L\forall$-property such that $\Gamma \subseteq \Delta$.*

PROOF: Assume all the formulas in $\mathcal{L}$ are enumerated. Define a sequence of lists of formulas $\Delta_0, \Delta_1, \ldots$ as follows: $\Delta_0 = \Gamma$. Given Δ_n, let α be the $n + 1$st formula in $\mathcal{L}$, and let $\Delta_{n+1} = \Delta_n \cup \{\alpha\}$ if $\Delta_n \cup \{\alpha\}$ is *AWARE*-consistent, or $\Delta_{n+1} = \Delta_n \cup \{\neg\alpha\}$ if not. In either case Δ_{n+1} is *AWARE*-consistent if Δ_n is. We can now let Δ be the union of all Γ_n. ■

The construction of the canonical structure is as follows. $\mathcal{W}$ is the set of all maximal *AWARE*-consistent lists of formulas with the $L\forall$-property. D is the set of all variables in $\mathcal{L}$; or equivalently, $D = X$. For every state w, which by construction is a list of formulas,

- D_w is the set of all variables x such that $E(x) \in w$;
- $\mathcal{P}_i(w)$ is the set of all states w' such that $L_i^-(w) \subseteq w'$, where $L_i^-(w)$ is the set of all formulas α such that $L_i\alpha \in w$;
- $\mathcal{A}_i(w)$ is the set of all variables x such that $A_iE(x) \in w$; and

- for every k-ary predicate P, $\pi(w)(P)$ is the set of all k-tuples $(x_1,\ldots,x_k)$ such that $P(x_1,\ldots,x_k) \in w$.

Notice that, for any list $w \in \mathcal{W}$, since w satisfies the part 1 of the $L_i\forall$-property, there must be at least one variable y such that $E(y) \in w$, and hence D_w is non-empty. Therefore the canonical structure is indeed an instance of the object-based unawareness structures.

Lemma 5 *If Γ is a maximal AWARE-consistent list of formulas with the $L\forall$-property, and α is a formula such that $L_i\alpha \notin \Gamma$, then there is a maximal AWARE-consistent list Δ of formulas with the $L\forall$-property such that $L_i^-(\Gamma) \cup \{\neg\alpha\} \subseteq \Delta$.*

Proof: Assume that all variables x are enumerated, and similarly for all formulas of the form $\forall x\theta$, and similarly for all formulas of the form $L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)$ with $h \geq 0$ and x not free in $x_1,\ldots,x_h$.

Define a sequence of formulas $\delta_0, \delta_1, \ldots$ as follows: δ_0 is $\neg\alpha$. Given δ_n, we define δ_{n+1} in two steps.

Step 1: We first extend δ_n to δ_n^+. Let $\forall x\theta$ be the $n+1$st formula of this form. Let y be the first variable such that $L_i^-(\Gamma) \cup \{\delta_n \wedge E(y) \wedge (\theta[y/x] \to \forall x\theta)\}$ is *AWARE*-consistent, and let δ_n^+ be $\delta_n \wedge E(y) \wedge (\theta[y/x] \to \forall x\theta)$.

We claim that, as long as $L_i^-(\Gamma) \cup \{\delta_n\}$ is *AWARE*-consistent, such a variable y must exist. Suppose not. Then for every variable y there is a finite sublist $\{L_i\beta_1,\ldots,L_i\beta_k\} \subset \Gamma$ such that $(\beta_1 \wedge \ldots \wedge \beta_k) \to \Big(E(y) \to \big(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\big)\Big)$ is a provable formula of *AWARE*. Therefore, by **LN** and **L**, the formula

$$(L_i\beta_1 \wedge \ldots \wedge L_i\beta_k) \to L_i\Big(E(y) \to \big(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\big)\Big)$$

is also a provable formula of *AWARE*. Since Γ is maximal *AWARE*-consistent and $L_i\beta_1,\ldots,L_i\beta_k \in \Gamma$, we have $L_i\Big(E(y) \to \big(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\big)\Big) \in \Gamma$ as well. And this is so for *every* variable y.

Since Γ has the $L\forall$-property, there is a variable y such that the formula

$$L_i\Big(E(y) \to \big(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\big)\Big) \to L_i\forall z\big(\delta_n \to \neg(\theta[z/x] \to \forall x\theta)\big)$$

is in Γ, where the variable z is chosen so that it does not occur in δ_n or in θ. Since $L_i\Big(E(y) \to \big(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\big)\Big)$ is in Γ for *every* variable y, the formula

$$L_i\forall z\big(\delta_n \to \neg(\theta[z/x] \to \forall x\theta)\big)$$

is in Γ. But z does not occur in δ_n or θ, and so by **E3** and **E4**, the formula

$$L_i\big(\delta_n \to \forall z\neg(\theta[z/x] \to \forall x\theta)\big)$$

is also in Γ.

However, by Lemma 3, the formula

$$\exists z(\theta[z/x] \rightarrow \forall x\theta)$$

is a provable formula of *AWARE*. So the formula

$$L_i\neg\delta_n$$

must also be a provable formula of *AWARE*, and hence is in Γ, or equivalently, $\neg\delta_n \in L_i^-(\Gamma)$. And this would make $L_i^-(\Gamma) \cup \{\delta_n\}$ *AWARE*-inconsistent, a contradiction.

Step 2: We next extend δ_n^+ to δ_{n+1}. Let $L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)$ be the $n{+}1$st formula of this form. We may assume that x is not free in δ_n^+ or in $\xi_1, \ldots, \xi_h$ since if it is we may choose a bound alphabetic variant of $\forall x\theta$ in which the variable that replaces x is not free in these formulas. Let y be the first variable such that $L_i^-(\Gamma) \cup \{\delta_n^+ \wedge \Bigg(L_i\Big(\xi_1 \rightarrow \cdots \rightarrow L_i\Big(\xi_h \rightarrow L_i\big(E(y) \rightarrow \theta[y/x]\big)\Big)\cdots\Big) \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)\Bigg)\}$ is *AWARE*-consistent, and let δ_{n+1} be $\delta_n^+ \wedge \Bigg(L_i\Big(\xi_1 \rightarrow \cdots \rightarrow L_i\Big(\xi_h \rightarrow L_i\big(E(y) \rightarrow \theta[y/x]\big)\Big)\cdots\Big) \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)\Bigg)$.

We claim that, as long as $L_i^-(\Gamma) \cup \{\delta_n^+\}$ is *AWARE*-consistent, such a variable y must exist. Suppose not. Then for every variable y there is a finite sublist $\{L_i\beta_1, \ldots, L_i\beta_k\} \subset \Gamma$ such that $(\beta_1 \wedge \ldots \wedge \beta_k) \rightarrow \Bigg(\delta_n^+ \rightarrow \neg\Big(L_i\Big(\xi_1 \rightarrow \cdots \rightarrow L_i\big(\xi_h \rightarrow L_i(E(y) \rightarrow \theta[y/x])\big)\cdots\Big) \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)\Big)\Bigg)$ is a provable formula of *AWARE*. Therefore, both

$$(\beta_1 \wedge \ldots \wedge \beta_k) \rightarrow \Bigg(\delta_n^+ \rightarrow L_i\Bigg(\xi_1 \rightarrow \cdots \rightarrow L_i\Big(\xi_h \rightarrow L_i\big(E(y) \rightarrow \theta[y/x]\big)\Big)\cdots\Bigg)\Bigg) \qquad (6)$$

and

$$(\beta_1 \wedge \ldots \wedge \beta_k) \rightarrow \Big(\delta_n^+ \rightarrow \neg L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i\forall x\theta)\cdots\big)\Big) \qquad (7)$$

are provable formulas of *AWARE*. From (6), by **LN** and **L**,

$$(L_i\beta_1 \wedge \ldots \wedge L_i\beta_k) \rightarrow L_i\Bigg(\delta_n^+ \rightarrow L_i\Bigg(\xi_1 \rightarrow \cdots \rightarrow L_i\Big(\xi_h \rightarrow L_i\big(E(y) \rightarrow \theta[y/x]\big)\Big)\cdots\Bigg)\Bigg) \quad (8)$$

is also a provable formula of *AWARE*. Since formulas $L_i\beta_1, \ldots, L_i\beta_k$ are all in Γ, so, from

(8), the formula

$$L_i\Bigg(\delta_n^+ \to L_i\Bigg(\xi_1 \to \cdots \to L_i\Big(\xi_h \to L_i\big(E(y) \to \theta[y/x]\big)\Big)\cdots\Bigg)\Bigg)$$

is also in Γ. And this is true for *every* variable y.

Since Γ has the $L_i\forall$ property, by a similar argument as in Step 1, the formula

$$L_i\Big(\delta_n^+ \to L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)\Big)$$

is also in Γ, or equivalently, the formula

$$\delta_n^+ \to L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)$$

is in $L_i^-(\Gamma)$. This, together with (7), would make $L_i^-(\Gamma) \cup \{\delta_n^+\}$ *AWARE*-inconsistent, a contradiction.

Since, by [cite a lemma here], $L_i^-(\Gamma) \cup \{\delta_0\}$ is *AWARE*-consistent, $L_i^-(\Gamma) \cup \{\delta_n^+\}$ and $L_i^-(\Gamma) \cup \{\delta_{n+1}\}$ are *AWARE*-consistent for all n.

Let Δ^- be the union of $L_i^-(\Gamma)$ and all the δ_n's. Δ^- is *AWARE*-consistent, and by construction, also possess the $L_i\forall$ property. Therefore, by Lemma 4, Δ^- can be extended into a maximal *AWARE*-consistent list Δ with the $L_i\forall$ property such that $L_i^-(\Gamma)\cup\{\neg\alpha\} \in \Delta$. ■

Lemma 6 *Let M be the canonical structure, and V be the valuation such that $V(x) = x$ for every variable/object $x \in X = D$. Then, for every maximal AWARE-consistent list $w \in \mathcal{W}$ of formulas with the $L\forall$-property, and for every formula $\alpha \in \mathcal{L}$, $(M, w, V) \models \alpha$ iff $\alpha \in w$.*

Proof: The proof proceeds by induction on the length of the formulas. For any atomic formula of the form $E(x)$, $(M, w, V) \models E(x)$ iff $V(x) \in D_w$ iff $x \in D_w$ iff $E(x) \in w$.

For any other atomic formula of the form $P(x_1, \ldots, x_k)$, $(M, w, V) \models P(x_1, \ldots, x_k)$ iff $\big(V(x_1), \ldots, V(x_k)\big) \in \pi(w)(P)$ iff $(x_1, \ldots, x_k) \in \pi(w)(P)$ iff $P(x_1, \ldots, x_k) \in w$.

For any formula of the form $\neg\alpha$, $(M, w, V) \models \neg\alpha$ iff $(M, w, V) \not\models \alpha$ which, by the induction hypothesis, is true iff $\alpha \notin w$ which, by the maximality of the list w, is true iff $\neg\alpha \in w$.

For any formula of the form $\alpha\wedge\beta$, $(M, w, V) \models \alpha\wedge\beta$ iff $(M, w, V) \models \alpha$ and $(M, w, V) \models \beta$ which, by the induction hypothesis, are true iff $\alpha \in w$ and $\beta \in w$ which, by the maximal *AWARE*-consistency of the list w, are true iff $\alpha \wedge \beta \in w$.

For any formula of the form $\forall x\alpha$, suppose $\forall x\alpha \in w$. Consider any x-alternative V' of V such that $V'(x) = y \in D_w$. Since $y \in D_w$, we have $E(y) \in w$. By **E2** and the maximal *AWARE*-consistency of the list w, we have $\alpha[y/x] \in w$. By the induction hypothesis, we have $(M, w, V) \models \alpha[y/x]$, which in turn implies $(M, w, V') \models \alpha$. Since this is true for every x-alternative V' of V such that $V'(x) \in D_w$, we have $(M, w, V) \models \forall x\alpha$.

Conversely, suppose $\forall x\alpha \notin w$. Since the list w possesses the $L\forall$-property, there is some variable y such that $E(y) \wedge (\alpha[y/x] \rightarrow \forall x\alpha) \in w$. By the maximal *AWARE*-consistency of the list w, we have $E(y) \in w$ (making $y \in D_w$) and $\alpha[y/x] \notin w$. By the induction hypothesis, the latter implies that $(M, w, V) \not\models \alpha[y/x]$, which in turn implies $(M, w, V') \not\models \alpha$, where V' is the x-alternative of V such that $V'(x) = y$. But $V'(x) \in D_w$, and hence we have $(M, w, V) \not\models \forall x\alpha$.

For any formula of the form $A_i\alpha$, let $\{x_1, \ldots, x_k\}$ be the free variables in α. If $k = 0$, then we have $(M, w, V) \models A_i\alpha$ and, by **A1** and the maximal *AWARE*-consistency of the list w, $A_i\alpha \in w$ as well. So let's assume $k \geq 1$. Since $V(x) \in \mathcal{A}_i(w)$ iff $x \in \mathcal{A}_i(w)$ iff $A_iE(x) \in w$, we have $(\mathcal{M}, w, V) \models A_i\alpha$ iff $A_iE(x) \in w$ for every $x \in \{x_1, \ldots, x_k\}$. By **A2**, **A3**, and the maximal *AWARE*-consistency of the list w, we have $A_iE(x) \in w$ for every $x \in \{x_1, \ldots, x_k\}$ iff $A_i\big(E(x_1) \wedge \ldots \wedge E(x_k)\big) \in w$ iff $A_i\alpha \in w$.

For any formula of the form $L_i\alpha$, suppose $L_i\alpha \in w$. Then we have $\alpha \in L_i^-(w)$, which implies $\alpha \in w'$ for every $w' \in \mathcal{P}_i(w)$. By the induction hypothesis, we have $(M, w', V) \models \alpha$ for every $w' \in \mathcal{P}_i(w)$, which implies $(M, w, V) \models L_i\alpha$.

Conversely, suppose $L_i\alpha \notin w$. By Lemma 5, there is an $w' \in \mathcal{W}$ such that $L_i^-(w) \cup \{\neg\alpha\} \subseteq w'$. Since $L_i^-(w) \subseteq w'$, we have $w' \in \mathcal{P}(w)$. Since $\neg\alpha \in w'$, by the induction hypothesis, we have $(M, w', V) \not\models \alpha$. Combining the two, we have $(M, w, V) \not\models L_i\alpha$.

For any formula of the form $K_i\alpha$, $(M, w, V) \models K_i\alpha$ iff $(M, w, V) \models A_i\alpha$ and $(M, w, V) \models L_i\alpha$ which, by the induction hypothesis, are true iff $A_i\alpha \in w$ and $L_i\alpha \in w$ which, by **K** and the maximal *AWARE*-consistency of the list w, are true iff $K_i\alpha \in w$. ■

Proof of Theorem 1: That every provable formula of *AWARE* is valid in $\mathcal{M}$ follows from Lemma 1. To prove the converse, suppose formula $\alpha \in \mathcal{L}$ is not a provable formula of *AWARE*. Then $\neg\alpha$ is *AWARE*-consistent, and hence by Lemmas 2 and 4, there exists a maximal *AWARE*-consistent list $w \in \mathcal{W}$ with the $L\forall$-property that contains it. By Lemma 6, $(M, w, V) \models \neg\alpha$. Therefore α is not valid in the canonical structure M under the valuation V. Since the cononical structure is one instance of the object-based unawareness structures, this proves that α is not valid in $\mathcal{M}$. ■

References

Aumann, R. J. (1976): "Agreeing to Disagree," *Annals of Statistics*, 4, 1236–1239.

Chung, K.-S., and L. Fortnow (2007): "Loopholes," mimeo, University of Hong Kong and University of Chicago.

Dekel, E., B. L. Lipman, and A. Rustichini (1998): "Standard State-Space Models Preclude Unawareness," *Econometrica*, 66(1), 159–173.

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

Feinberg, Y. (2004): "Subjective Reasoning—Games with Unawareness," Research Paper No. 1875, Stanford Graduate School of Business.

Halpern, J. (2001): "Alternative Semantics for Unawareness," *Games and Economic Behavior*, 37(2), 321–339.

Halpern, J., and L. C. Rego (2006): "Reasoning About Knowledge of Unawareness," in *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning.*

Heifetz, A., M. Meier, and B. Schipper (2006): "Interactive Unawareness," *Journal of Economic Theory*, 130, 78–94.

——— (2007): "A Canonical Model of Interactive Unawareness," *Games and Economic Behavior.*

Hughes, G., and M. Cresswell (1996): *A New Introduction to Modal Logic*. Routledge, London and New York.

Li, J. (2006): "Informational Structures with Unawareness," mimeo, University of Pennsylvania.

Modica, S., and A. Rustichini (1994): "Awareness and Partitional Information Structures," *Theory and Decision*, 37, 107–124.

——— (1999): "Unawareness and Partitional Information Structures," *Games and Economic Behavior*, 27(2), 265–298.

Russell, B. (1912): *The Problems of Philosophy*. T. Butterworth, London.

Sillari, G. (2006): "Models of Unawareness," in *Logic and the Foundations of Game and Decision Theory: Proceedings of the Seventh Conference*, ed. by G. Bonanno, W. van der Hoek, and M. Woolridge.