OBJECT-BASED UNAWARENESS: AXIOMS

Oliver J. Board

Paul | Weiss, USA
ojboard@paulweiss.com

Kim-Sau Chung

Hong Kong Baptist University, China
kschung@hkbu.edu.hk

ABSTRACT

This paper provides foundations for a model of unawareness, called object-based unawareness (OBU) structures, that can be used to distinguish between what an agent is unaware of and what she simply does not know. At an informal level, this distinction plays a key role in a number of papers such as Tirole (2009) and Chung & Fortnow (2016). In this paper, we give the model-theoretic description of OBU structures by showing how they assign truth conditions to every sentence of the formal language used. We then prove a model-theoretic sound and completeness theorem, which characterizes OBU structures in terms of a system of axioms. We then verify that agents in OBU structures do not violate any of the introspection axioms that are generally considered to be necessary conditions for a plausible notion of unawareness. Applications are provided in our companion paper.

Keywords: Awareness, object-based unawareness, modal logic.

JEL Classification Numbers: D83, D91.

1. INTRODUCTION

THERE are two literatures on unawareness, and it is often not clear that the authors in each group are aware of each other's contributions.

The first unawareness literature (let's call it the *applied* literature) consists of applied models, such as Tirole (2009) and Chung & Fortnow (2016), where agents are uncertain whether they are aware of everything that their opponents are aware of, and have to strategically interact under these uncertainties. For example, in Tirole (2009), a buyer and a seller negotiate a contract as in the standard holdup problem. At the time of negotiation, there may or may not exist a better design for the product. Even if a better design exists, however, the contracting parties may not be aware of it. If a party is aware of it, he can choose whether or not to point it out to the other party. But even if he is not aware of it, he *is* aware that a better design may exist and his opponent may be aware of this better design. In Tirole's words, "parties are unaware, but aware that they are unaware"; and they have to negotiate under this uncertainty. Chung & Fortnow (2016) consider the plight of an American founding father drafting a Bill of Rights that will be interpreted by a judge 200 years later. The founding father is aware of some human rights, but is uncertain whether or not there are other human rights that he is unaware of. Here, as in Tirole (2009), the founding father is unaware, but aware that he may be unaware; and he has to choose how he should write the Bill of Rights in the face of this uncertainty.

The second unawareness literature (let's call it the *foundational* literature) attempts to provide a more rigorous account of the properties of unawareness: see e.g. Fagin & Halpern (1988), Modica & Rustichini (1994), Modica & Rustichini (1999), Dekel et al. (2016), Halpern (2001), Li (2006), Halpern & Rego (2006), Sillari (2006), and Heifetz et al. (2006b), Heifetz et al. (2006a). These authors are motivated by the concern that *ad hoc* applied models, if not set up carefully enough, may go awry in the sense that agents in those models may violate rationality in some way, as captured by various introspection axioms (which we shall refer to as the DLR axioms hereafter).[1] This concern is articulated in Modica & Rustichini (1994), and Dekel et al. (2016). The rest of this literature proposes various models that are set up carefully enough to

---

[1] In particular, two of the key axioms that lie behind Dekel et al.'s (2016) impossibility result are *KU-introspection* ("the agent cannot know that he is unaware of a specific event") and *AU-introspection* ("if an agent is unaware of an event E, then he must be unaware of being unaware of E").

take these concerns into account.

These two literatures are somewhat disconnected. For example, Tirole makes no reference to any work in the foundational literature, nor does he explain whether or not his agents satisfy the introspection axioms that are one of the main concerns of that literature. Similarly, none of the papers in the foundational literature explain whether Tirole's model may fit in their framework, and if not, whether Tirole's agents violate some or all of the introspection axioms. This paper and its companion paper, Board & Chung (2008), attempt to connect these two literatures.

There is a reason why it is difficult to directly compare Tirole's model with the majority of the models proposed in the foundational literature. To propose a model, and to provide foundations for it, an author needs to explain how her model should be interpreted. In several of the papers discussed above (e.g. Fagin & Halpern (1988)), this is done by showing how a formal structure assigns truth conditions to each sentence of some formal language; i.e., by the procedure of systematically giving yes/no answers to a laundry list of questions such as: "At state w, does agent i know that it is sunny in New York?" In many of the papers more familiar to economists (e.g. Li (2006)), although this procedure is not performed explicitly, there is typically a clear way to assign truth conditions to an appropriately-specified language according to the author's description of her model. But the procedure of assigning truth-conditions is well-defined only if the set of questions (to be given yes/no answers) is defined clearly. This set of questions is determined by the language associated with the proposed model, and is chosen (either explicitly or implicitly) by the author. But this also means that we can understand a proposed model only up to its associated language. If we ask a question that does not belong to the associated language, we cannot expect to find an answer.

Unfortunately, questions such as "At state w, does agent i know that he is not aware of everything?" do *not* belong to the language of many of the studies in foundational literature (notable exceptions include Halpern & Rego (2006) and Sillari (2006), which we return to in the next paragraph). More generally, the languages underlying many of the studies cited above do not contain quantifiers; while sentences such as "agent i is aware of everything" (implicitly) do. This provides some explanation as to why it is difficult to compare the two literatures. In other words, while in Tirole's model, "parties are unaware, but aware that they are unaware", it is difficult to figure out when or if this would be true of the agents in most of the models proposed in the

foundational literature. Those models do not address such questions, and hence our understanding of them is somewhat limited.

Several contributions by logicians and computer scientists, however, present models that *do* address these questions (e.g., Halpern & Rego (2006) and Sillari (2006)). These papers explicitly present and analyze formal languages that contain quantifiers, and are thus richer than the languages underlying the models discussed above. Their models, however, are very different from the applied models used by Tirole (2009) and Chung & Fortnow (2016). For example, Halpern & Rego (2006) describe an agent's awareness by means of a *syntactic* awareness function mapping states to sets of sentences of the formal language, to be interpreted as the set of sentences the agent is aware of. Certain restrictions are then imposed on the form of this function to capture a plausible notion of awareness. This "list of sentences" approach is more general, but the cost of this additional generality is less structure. This may explain why this approach, while not uncommon in the formal logic literature, is rarely seen in economics.[2]

In the specific case of Halpern & Rego (2006), there is a more specific reason why it could not be used to provide foundations for applied models such as Tirole (2009) and Chung & Fortnow (2016). In both of these papers, although agents know what they are aware of, they may be uncertain whether or not they are aware of everything. Such uncertainty cannot arise in Halpern & Rego (2006), however.[3] To capture this kind of uncertainty, they would have

---

[2] To provide an analogy that may help elucidate this comparison, consider the difference between Aumann's information partition model, where a partition of the state space is used to derive an agent's knowledge of events, and a "list of sentences" approach where knowledge is instead modeled by a list of sentences describing exactly what that agent knows.

[3] For readers who are familiar with Halpern & Rego (2006), this can be proved formally as follows. Recall the following definition in Halpern & Rego (2006): "*Agents know what they are aware of* if, for all agents i and all states s, t such that $(s, t) \in \mathcal{K}_i$ we have that $\mathcal{A}_i(s) = \mathcal{A}_i(t)$." So it suffices to prove that, in any instance of Halpern & Rego (2006) structure, if there is a state t such that agent i is uncertain whether or not there is something he is unaware of, then there must be another state s such that $(s, t) \in \mathcal{K}_i$ but $\mathcal{A}_i(s) \neq \mathcal{A}_i(t)$. Let $\alpha = \exists x \neg A_i x$ represent "there is something that agent i is unaware of". Therefore, $\neg \alpha$ means "there is nothing that agent i is unaware of". Let $\beta = A_i \alpha \wedge A_i \neg \alpha \wedge \neg X_i \alpha \wedge \neg X_i \neg \alpha$ represent "agent i is aware of both α and $\neg \alpha$ but he does not know whether α or $\neg \alpha$ is true (recall that X_i is Halpern & Rego (2006) explicit knowledge operator). In short, β means "agent i is uncertain whether or not there is something he is unaware of". Let M be any instance of Halpern & Rego (2006) structure, and t is a state such that $(M, t) \models \beta$. Then we have $(M, t) \models \neg K_i \alpha \wedge \neg K_i \neg \alpha$ (recall that K_i is Halpern & Rego (2006) implicit knowledge operator). Therefore, there exists

to consider a framework in which the formal language is allowed to vary across states: an agent who is unable to distinguish between two states with different languages could thus be uncertain about how many sentences there are, and hence uncertain about how many she is unaware of.

To summarize, while the assumption that "agents are unaware, but are aware that they are unaware" plays a key role in much of the applied literature on unawareness, the foundations of these models remain unclear. We do not know whether agents in these models violate some or all the introspection axioms that are one of the main concerns of the foundational literature. This paper and its companion paper, Board & Chung (2008), attempt to provide this missing foundation.

In these two papers, we describe a model, or more precisely a class of models, called *object-based unawareness structures* (OBU structures). Readers will find that these structures encompass models used in the applied literature. In comparison with the applied literature, however, we provide complete and rigorous foundations for these structures. The underlying language we use is rich, and in particular contains quantifiers, enabling us to describe explicitly whether or not agents are aware that they are unaware. We will provide an axiomatization for these structures, verifying that all of the appropriate introspection axioms are satisfied. The value of thinking about agents who exhibit this kind of uncertainty has already been demonstrated by the existing applied literature; we demonstrate the tractability of our framework by considering further applications.

A key feature of our structures is that unawareness is object-based:[4] a seller may be unaware of a better design, or a founding father may be unaware of a particular human right. In contrast, in models of unforeseen contingencies, agents are unaware of contingencies, or states. This raises the question of whether the agents in our model are aware of every state. We do not have answer to this question. As we argued above, our understanding of any given model is constrained by the language we choose to work with. Although our language is one of the richest in the foundational literature, there are questions that fall outside of it. We do not have answers to these questions, simply because we do not speak that language.

---

a state s such that $(t, s) \in \mathcal{K}_i$ and $(M, s) \models \neg\alpha$, and another state s' such that $(t, s') \in \mathcal{K}_i$, and $(M, s') \models \alpha$. Since $\alpha = \exists x \neg A_i x$, there exists ϕ such that $\phi \in \mathcal{A}_i(s)$ and $\phi \notin \mathcal{A}_i(s')$. But that means at least of one $\mathcal{A}_i(s)$ and $\mathcal{A}_i(s')$ is different from $\mathcal{A}_i(t)$.

[4] We discuss other possible sources of unawareness in the conclusion.

The division of labor between this paper and its companion paper, Board & Chung (2008), is as follows. In this paper, we give the model-theoretic description of OBU structures by showing how they assign truth conditions to every sentence of the formal language. We then prove a model-theoretic sound and completeness theorem, which characterizes OBU structures in terms of a system of axioms. We then verify that agents in OBU structures do not violate any of the introspection axioms that are generally considered to be necessary conditions for a plausible notion of unawareness. This paper also contains a more complete literature review, as well as a discussion of several variants of OBU structures.

In our companion paper, Board & Chung (2008), we give a set-theoretic description of the OBU structures. Although less formal than the model-theoretic treatment, we hope this will be more accessible to the general audience. In parallel to the model-theoretic sound and completeness theorem in this paper, we prove set-theoretic completeness results in Board & Chung (2008).

The second half of Board & Chung (2008) considers two applications. First, we use the model to provide a justification for the *contra proferentem* doctrine of contract interpretation, commonly used to adjudicate ambiguities in insurance contracts. Under *contra proferentem*, ambiguous terms in a contract are construed against the drafter. Our main result is that when drafter (the insurer) has greater awareness than the other party (the insured), *and when the insured is aware of this asymmetry*, *contra proferentem* minimizes the chances that the insured forgoes gain of trade for fear of being exploited. On the other hand, when there is no asymmetric awareness, efficiency considerations suggest no reason to prefer *contra proferentem* over an alternative interpretive doctrine that resolves ambiguity in favor of the drafter.

From the perspective of our framework, an argument common among legal scholars as far back as Francis Bacon, that *contra proferentem* encourages the insurer to write clearer contracts, misses the point. If a more precise contract increases the surplus to be shared between the insurer and the insured, market forces provide incentives to draft such a contract regardless of the interpretive doctrine employed by the court. The advantage of *contra proferentem* is rather that it enables the insurer to draft more acceptable contracts, by expanding the set of events that he can credibly insure.

Our second application examines speculative trade. We first generalize the classical No Trade Theorem to situations where agents are delusional

but nevertheless satisfy a weaker condition called terminal partitionality. We then introduce the concepts of *living in denial* (i.e., agents believe, perhaps incorrectly, that there is nothing that they are unaware of) and *living in paranoia* (i.e., agents believe, perhaps incorrectly, that there is something that they are unaware of). We show that both living in denial and living in paranoid, in the absence of other forms of delusion, imply terminal partitionality, and hence the no trade theorem result obtains.

The structure of this paper is as follows: 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 the DLR 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

We start by introducing our language. Formally, a *language* is a set of sentences; it should be rich enough to express everything we might want to say about the agents in our model, such as "the founding father knows that there are some human rights he is not aware of". As we explain below, this requires working with a version of first-order modal logic rather than propositional modal logic, which (implicitly or explicitly) forms the basis of the majority of the models developed in the preceding literature on unawareness.

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 $\mathcal{M}$. 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 be 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 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 way, modal operators allow us to construct formulas that describe the mind of an agent. 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 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 entered 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:

- if $\phi \in \Phi$, then $\phi \in \mathcal{L}$;

- if $\phi, \psi \in \mathcal{L}$, then $\neg\phi \in \mathcal{L}$ and $\phi \wedge \psi \in \mathcal{L}$;

- if $\phi \in \mathcal{L}$ and $x \in X$, then $\forall x\phi \in \mathcal{L}$;

- if $\phi \in \mathcal{L}$ and $i \in N$, then $L_i\phi \in \mathcal{L}$ and $A_i\alpha \in \mathcal{L}$ and $K_i\alpha \in \mathcal{L}$.

---

[5] Augmenting propositional logic (rather than first-order logic) with modal operators generates *propositional modal logic*: this is the language used in most of the previous literature on unawareness. It should be clear that propositional modal logic does not include formulas such as "I am not sure whether or not you are aware of something that I am not".

[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.

We use the following standard abbreviations:

- $\alpha \lor \beta$ for $\neg(\neg\alpha \land \neg\beta)$;

- $\alpha \to \beta$ for $\neg\alpha \lor \beta$;

- $\alpha \leftrightarrow \beta$ for $(\alpha \to \beta) \land (\beta \to \alpha)$;

- $\exists x\alpha$ 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 list of human 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 $\forall x$.[7] 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:

---

[7] 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 the 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 $\alpha \land \beta$ 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.

PC all propositional tautologies are axioms.

MP from α and $\alpha \to \beta$ infer β.

PC (propositional calculus) is a set of axioms: propositional tautologies include formulas such as $\alpha \vee \neg\alpha$ and $(Tall(x) \wedge \exists x \, Taller(x, y)) \to \exists x \, Taller(x, y)$. **MP** (*modus ponens*) is an inference rule, and says that if α and $\alpha \to \beta$ 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, $\forall x \alpha \to (E(y) \to \alpha[y/x])$ is an axiom.

E3 for any formulas α and β and variable x, $\forall x(\alpha \to \beta) \to (\forall x \alpha \to \forall x \beta)$ 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 $\forall x \alpha$.

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 & Rego (2006) and Sillari (2006).

In axiom **E2**, $\alpha[y/x]$ is the same formula as α with free y replacing every free x.[8] 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 then 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 of 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.

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.

---

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

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, 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 $\alpha_1 \rightarrow L_i(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\beta)\cdots)$, where $h \geq 0$, infer $\alpha_1 \rightarrow L_i(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\forall x\beta)\cdots)$, 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 bit of a 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 (Theorem 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

$$M = (W, D, \mathcal{E}, \mathcal{P}_1, \ldots, \mathcal{P}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi),$$

where:

- W is a set of possible worlds, with typical element w;

- D is a set of objects;

- $\mathcal{E} : W \to D$ is an *existence correspondence*: $\mathcal{E}(w)$ is the set of objects which are real at world w;

- $\mathcal{P}_i : W \to 2^W$ is agent i's *possibility correspondence*: $\mathcal{P}_i(w)$ is the set of worlds that agent i considers possible when the true world is w;

- $\mathcal{A}_i : W \to 2^D$ is agent i's *awareness correspondence*: $\mathcal{A}_i(w)$ is the set of objects that agent i is aware of when the true world is 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 $E := \{w \mid |John \in \pi(w)(Tall)\}$. We then ask two questions: in world w, (1) does i implicitly know that John is tall ($\mathcal{P}(w) \subset E$)? 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 $E := \{w \mid D_w \subset \pi(w)(tall)\}$. 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 $(V(x_1), \ldots, V(x_k)) \in \pi(w)(P)$;

$(M, w, V) \models \neg\alpha$ 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 $V'(x) \in D_w$;[9]

$(M, w, V) \models A_i\alpha$ 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 $w' \in \mathcal{P}(w)$;

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

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

$$C \models \alpha.$$

2.4. The Characterization Theorem

Theorem 1. *$\phi \in \mathcal{L}$ is valid in $\mathcal{M}$ if and only if ϕ is provable 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.[10] We present the complete proof in the appendix.

---

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

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

3. AN APPLICATION

Chung & Fortnow (2016) 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 elaborate the Bill is. The more elaborate the Bill is, the more likely the judge will be to rule that it is constitutional for the government to infringe those rights not included into the Bill.

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 & Fortnow (2016) 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 arbitrary property that both objects satisfy in both worlds (i.e., $\pi(w)(P) = \{s, f\}$ for $w = w_1, w_2$), then we have:

$$(M, w_1) \models \exists x U_i P(x),$$
$$(M, w_2) \models \neg \exists x U_i P(x).$$

Suppose, in the true state w_2, the agent cannot distinguish w_1 and w_2 (i.e., $\mathcal{P}_i(w_2) = \{w_1, w_2\}$). 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:

$$(M, w_2) \models A_i \left(\neg \exists x U_i P \left(x \right) \right), \quad \text{and}$$
$$(M, w_2) \models A_i \left(\exists x U_i P \left(x \right) \right).$$

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 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 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 & Rego (2006) and Sillari (2006), where they adopt the possibilitist existence axioms, their structures are necessarily characterized by constant domains instead.

4. THE DLR AXIOMS

A first impression of some readers of this paper is that object-based unawareness structures violate Dekel et al.'s (2016) AU Introspection Axiom. AU Introspection is represented by formulas of the form $U_i \phi \rightarrow U_i U_i \phi$. Indeed, every such formula is a provable in *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 $A_i U_i \phi \rightarrow A_i \phi$ is a provable formula of *AWARE* (for all $\phi \in \mathcal{L}$ and all i). By simple propositional reasoning, then, $U_i \alpha \rightarrow U_i U_i \alpha$ is also a provable formula of *AWARE*.

Consider however a similar formula: $U_i\alpha \rightarrow U_i\exists x U_i\alpha$. 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**.

Dekel et al. (2016) posit a third axiom of 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.[11] In terms of our structures, **T** corresponds to the restriction that the possibility correspondences are reflexive: $w \in \mathcal{P}(w)$. 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**.

The impossibility result of Dekel et al. (2016) 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

---

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

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 V, if the formula α is true (i.e., $(M, w, V) \models \alpha$) in every world w then the formula $K_i \alpha$ is also true (i.e., $(M, w, V) \models K_i \alpha$) in every world w. Monotonicity corresponds to the restriction that, for any given OBU structure M and valuation V, if the formula $\alpha \to \beta$ is true (i.e., $(M, w, V) \models \alpha \to \beta$) in every world w then the formula $K_i \alpha \to K_i \beta$ is also true (i.e., $(M, w, V) \models K_i \alpha \to K_i \beta$) 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 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	$L_i \phi \to \phi$
PI	$L_i \phi \to L_i L_i \phi$
NI	$\neg L_i \phi \to L_i \neg L_i \phi$

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 $w \in \mathcal{P}(w)$ for all w. We say that it is *transitive* if $x \in \mathcal{P}(w)$ and $y \in \mathcal{P}(x)$ imply $y \in \mathcal{P}(w)$ for all w, x, y; and *Euclidean* if $x \in \mathcal{P}(w)$ and $y \in \mathcal{P}(w)$ imply $y \in \mathcal{P}(x)$ 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 Euclideanness to **NI**:

Theorem 2. *The set of formulas $\alpha \in \mathcal{L}$ that are valid in $\{\mathcal{M}^r, \mathcal{M}^t, \mathcal{M}^e, \mathcal{M}^{rt},$ $\mathcal{M}^{re}, \mathcal{M}^{re}, \mathcal{M}^{rte}\}$ 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 $A_i\phi \to K_iA_i\phi$.

A related axiom, A-Introspection ($A_i\alpha \leftrightarrow K_iA_i\alpha$), appears in Heifetz et al. (2006b). 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 $A_i\phi \to L_iA_i\phi$ is an axiom.

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

LA2 $U_i\phi \to L_iU_i\phi$ is an axiom.

LA2 has appeared in some earlier studies on unawareness.[12] We cannot, however, think of any justification for it,[13] 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 $w' \in \mathcal{P}_i(w)$, $\mathcal{A}_i(w) = \mathcal{A}_i(w')$.

---

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

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

6. LITERATURE REVIEW

In this section we shall discuss some of the important contributions to the literature on unawareness.[14]   All of these papers share a common feature: unawareness is associated with events/facts instead of with objects/things.

In an early paper, Fagin & 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[15] 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 & Rustichini (1999) provide the first treatment of unawareness in the economics literature that avoids the Dekel et al. (2016) critique (and also address concerns raised in an earlier paper of their own, Modica & 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 & 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.

Heifetz et al. (2006b) 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

---

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

[15] 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.

the awareness of agents. In a companion paper, Heifetz et al. (2006a) 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 sentences such as "I am not sure whether or not you are aware of something that I am not". Halpern & Rego (2006) use second-order modal logic, augmenting the language of Fagin & 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*,[16] 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 $\mathcal{E}(w) = W$ for every $w \in W$. If we further add Axioms **LA1** and **LA2**—which, as we explained in

---

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

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 0 and 1.

As Sillari (2006) points out, this very problem arises in Halpern & 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 & Rego (2006) 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.

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 & 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 & Halpern (1988) by introducing an unawareness function that assigns to each agent in 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 W, D, $\mathcal{E}$, $\mathcal{P}_i$, and $\mathcal{A}_i$) as 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 & 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, in 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 in 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

In this paper we have proposed a user-friendly model that allows us to express 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 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 explicit 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 used to model the actions of 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 every time. 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 3. *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 $(M, w, V) \models \forall x(\alpha \rightarrow \beta)$ 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 $(M, w, V') \models \alpha \rightarrow \beta$ 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 $(M, w, V) \models L_i(\alpha \rightarrow \beta)$ 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 $(M, w', V) \models \alpha \rightarrow \beta$ 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 $(M, w', V) \not\models \alpha$, which implies the formula α is not valid in M.

For **UGL**, suppose $(M, w, V) \not\models \alpha_1 \rightarrow L_i(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\forall x\beta)\cdots)$. 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 $(M, w, V') \not\models \alpha_1 \rightarrow L_i(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\beta)\cdots)$, which in turn implies the formula $\alpha_1 \rightarrow L_i(\alpha_2 \rightarrow \cdots \rightarrow L_i(\alpha_h \rightarrow L_i\beta)\cdots)$ is not valid in M. $\square$

By the truth-value-assignment rule in Section 2.3, the formula $\alpha \wedge \neg\alpha$ is not valid in M, and hence by Lemma 3 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 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 $E(y) \wedge (\alpha[y/x] \rightarrow \forall x\alpha)$ 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
$$L_i\left(\beta_1 \rightarrow \cdots \rightarrow L_i\left(\beta_h \rightarrow L_i\big(E(y) \rightarrow \alpha[y/x]\big)\right) \cdots\right) \rightarrow L_i\big(\beta_1 \rightarrow$$
$$\cdots \rightarrow L_i(\beta_h \rightarrow L_i\forall x\alpha) \cdots\big) \text{ is in } \Gamma.$$

Lemma 4. *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 4, we need another lemma first.

Lemma 5. *The formula $\exists y(\theta[y/x] \rightarrow \forall x\theta)$ 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 $\forall xE(x) \rightarrow \forall x(\forall y\theta[y/x] \rightarrow \theta)$ 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 4: 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 $L_i\big(\xi_i \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)$ 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

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

We claim that, as long as Γ_n is *AWARE*-consistent, Γ_n^+ is *AWARE*-consistent as well. Suppose not. Then the formula $\beta \to \big(E(y) \to \neg(\theta[y/x] \to \forall x\theta)\big)$, 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 $\beta \to \big(\forall y E(y) \to \forall y\neg(\theta[y/x] \to \forall x\theta)\big)$ is a provable formula. By **E1**, the formula $\beta \to \forall y\neg(\theta[y/x] \to \forall x\theta)$ is a provable formula. By Lemma 5, the formula $\neg\beta$ is a "theroem," contradicting the presumption that Γ_n is *AWARE*-consistent.

Step 2: We next extend Γ_n^+ to Γ_{n+1}. Let $L_i\big(\xi_i \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)$ 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 $\Gamma_{n+1} = \Gamma_n^+ \cup \{L_i\big(\xi_1 \to \cdots \to L_i\big(\xi_h \to L_i\big(E(y) \to \theta[y/x]\big)\big)\cdots\big) \to L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)\}$.

We claim that, as long as Γ_n^+ is *AWARE*-consistent, Γ_{n+1} is *AWARE*-consistent as well. Suppose not. Then both formulas

$$\beta \to L_i\left(\xi_1 \to \cdots \to L_i\big(\xi_h \to L_i\big(E(y) \to \theta[y/x]\big)\big)\cdots\right) \qquad (3)$$

and

$$\beta \to \neg L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big), \qquad (4)$$

where β denotes 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

$$\beta \;\rightarrow\; L_i\left(\xi_1 \rightarrow \cdots \rightarrow L_i\left(\xi_h \rightarrow L_i \forall y\big(E(y) \rightarrow \theta[y/x]\big)\right)\cdots\right) \qquad (5)$$

is a provable formula.

By **E3**, **LN**, and **L**, the formula $L_i \forall y\big(E(y) \rightarrow \theta[y/x]\big) \rightarrow \big(L_i \forall y E(y) \rightarrow L_i \forall y \theta[y/x]\big)$ 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 $L_i \forall y\big(E(y) \rightarrow \theta[y/x]\big) \rightarrow L_i \forall y \theta[y/x]$ is a provable formula. By using **LN** and **L** repeatedly (for h times to be exact), we infer that the formula $L_i\left(\xi_1 \rightarrow \cdots \rightarrow L_i\left(\xi_h \rightarrow L_i \forall y\big(E(y) \rightarrow\right.\right.$

$\left.\left.\theta[y/x]\big)\right)\cdots\right) \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i \forall y \theta[y/x])\cdots\big)$ is a provable formula. From this, together with (5), we infer that $\beta \rightarrow L_i\big(\xi_1 \rightarrow \cdots \rightarrow L_i(\xi_h \rightarrow L_i \forall y \theta[y/x])\cdots\big)$ 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 6. *If an AWARE-consistent list Γ of formulas possesses the $L\forall$-property, then there is a maximal AWARE-consistent list Δ of 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 $A_i E(x) \in w$; 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 7. *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 $L_i(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta) \cdots)$ 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 $(\beta_1 \wedge \ldots \wedge \beta_k) \to \left(E(y) \to \left(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\right)\right)$ is a provable formula of *AWARE*. Therefore, by **LN** and **L**, the formula

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

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 $L_i\left(E(y) \to \left(\delta_n \to \neg(\theta[y/x] \to \forall x\theta)\right)\right) \in \Gamma$ 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

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

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 5, the formula

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

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 $L_i(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots)$ 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 $L_i^-(\Gamma) \cup \{\delta_n^+ \wedge \Big(L_i\Big(\xi_1 \to \cdots \to L_i\Big(\xi_h \to$

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

is *AWARE*-consistent, and let δ_{n+1} be $\delta_n^+ \wedge \Big(L_i\Big(\xi_1 \to \cdots \to L_i\Big(\xi_h \to$

$L_i\big(E(y) \to \theta[y/x]\big)\Big)\cdots\Big) \to L_i\big(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\big)\Big)$.

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 $(\beta_1 \wedge \ldots \wedge \beta_k) \to \Big(\delta_n^+ \to$

$$\neg\left(L_i\left(\xi_1 \to \cdots \to L_i\left(\xi_h \to L_i(E(y) \to \theta[y/x])\right)\cdots\right) \to L_i\left(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\right)\right)$$

is a provable formula of *AWARE*. Therefore, both

$$(\beta_1 \wedge \ldots \wedge \beta_k) \to \left(\delta_n^+ \to L_i\left(\xi_1 \to \cdots \to L_i\left(\xi_h \to L_i(E(y) \to \theta[y/x])\right)\cdots\right)\right) \tag{6}$$

and

$$(\beta_1 \wedge \ldots \wedge \beta_k) \to \left(\delta_n^+ \to \neg L_i\left(\xi_1 \to \cdots \to L_i(\xi_h \to L_i\forall x\theta)\cdots\right)\right) \tag{7}$$

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

$$(L_i\beta_1 \wedge \ldots \wedge L_i\beta_k) \to L_i\left(\delta_n^+ \to L_i\left(\xi_1 \to \cdots \to L_i\left(\xi_h \to L_i(E(y) \to \theta[y/x])\right)\cdots\right)\right) \tag{8}$$

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

$$L_i\left(\delta_n^+ \to L_i\left(\xi_1 \to \cdots \to L_i\left(\xi_h \to L_i(E(y) \to \theta[y/x])\right)\cdots\right)\right)$$

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

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

is also in Γ, or equivalently, the formula

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

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

Since $L_i^-(\Gamma) \cup \{\delta_0\}$ 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 6, 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 8. *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 $(V(x_1), \ldots, V(x_k)) \in \pi(w)(P)$ 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 $V(x) \in \mathscr{A}_i(w)$ iff $x \in \mathscr{A}_i(w)$ iff $A_i E(x) \in w$, we have $(M, w, V) \models A_i\alpha$ iff $A_i E(x) \in w$ for every $x \in \{x_1, \ldots, x_k\}$. By **A2**, **A3**, and the maximal *AWARE*-consistency of the list w, we have $A_i E(x) \in w$ 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 7, 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 $K_i\alpha \in w$. $\qquad\square$

PROOF OF THEOREM 1: That every provable formula of *AWARE* is valid in M follows from Lemma 3. 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 4 and 6, there exists a maximal *AWARE*-consistent list $w \in \mathcal{W}$ with the $L\forall$-property that contains it. By Lemma 8, $(M, w, V) \models \neg\alpha$. Therefore α is not valid in the canonical structure M under the valuation V. Since the canonical structure is one instance of the object-based unawareness structures, this proves that α is not valid in $\mathcal{M}$. $\qquad\square$

References

Aumann, R. (1976). Agreeing to disagree. *Annals of Statistics*, *4*, 1236–9.

Board, O. J., & Chung, K.-S. (2008). Object-based unawareness: Theory and applications. *University of Pittsburgh*, *Working Paper*.

Chung, K.-S., & Fortnow, L. (2016). Loopholes. *Economic Journal*, *126*, 1774–1797.

Dekel, E., Lipman, B., & Rustichini, A. (2016). Standard state space models preclude unawareness. *Econometrica*, *66*, 159–173.

Fagin, R., & Halpern, J. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, *34*, 39–76.

Feinberg, Y. (2004). Subjective reasoning. *Games with Unawareness Research Paper, Stanford Graduate School of Business.*, *1875*.

Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior*, *37*, 321–339.

Halpern, J. Y., & Rego, L. C. (2006). Reasoning about knowledge of unawareness. In *Proceedings of the tenth international conference on principles of knowledge representation and reasoning* (pp. 14–24).

Heifetz, A., Meier, M., & Schipper, B. C. (2006a). A canonical model of interactive unawareness. *The University of California, Davis*, *Working Paper*.

Heifetz, A., Meier, M., & Schipper, B. C. (2006b). Interactive unawareness. *Journal of Economic Theory*, *130*, 78–94.

Hughes, G. E., & Cresswell, M. J. (1996). *A New Introduction to Modal Logic*. Routledge, London.

Li, J. (2006). Informational structures with unawareness. *University of Pennsylvania*, *Working Paper*.

Modica, S., & Rustichini, A. (1994). Awareness and partitional information structures. *Theory and Decision*, *37*, 107–124.

Modica, S., & Rustichini, A. (1999). Unawareness and partitional information structures. *Games and Economic Behavior*, *27*, 265–298.

Sillari, G. (2006). Models of unawareness. In G. Bonanno, W. van der Hoek, & M. Wooldridge (Eds.), *Logic and the Foundations of Game and Decision Theory, Proceedings of the Seventh Conference* (pp. 209–40). Amsterdam University Press.

Tirole, J. (2009). Cognition and incomplete contracts. *American Economic Review*, *99*, 265–294.