Artificial Intelligence 175 (2011) 220–235

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

# Dealing with logical omniscience: Expressiveness and pragmatics ☆

Joseph Y. Halpern [a], Riccardo Pucella [b,*]

[a] Cornell University, Ithaca, NY 14853, USA
[b] Northeastern University, Boston, MA 02115, USA

ARTICLE INFO

*Article history:*
Available online 3 April 2010

*Keywords:*
Logic
Knowledge
Logical omniscience
Awareness
Impossible worlds
Probability

ABSTRACT

We examine four approaches for dealing with the logical omniscience problem and their potential applicability: the syntactic approach, awareness, algorithmic knowledge, and impossible possible worlds. Although in some settings these approaches are equi-expressive and can capture all epistemic states, in other settings of interest (especially with probability in the picture), we show that they are not equi-expressive. We then consider the pragmatics of dealing with logical omniscience—how to choose an approach and construct an appropriate model.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

John McCarthy was a pioneer in the use of reasoning about knowledge in AI. His notion of what "any fool" knows, one of the earliest uses of common knowledge, goes back to roughly 1970; it first appears in a published paper in [21]. It thus seems particularly appropriate for a paper on logics of knowledge to appear in this special issue of *Artificial Intelligence* dedicated to John McCarthy and his work.

Like most authors, McCarthy gave "possible-worlds" style semantics to knowledge. Logics of knowledge based on possible-worlds semantics have been shown to be useful in many areas of knowledge representation and reasoning, ranging from security to distributed computing to game theory. In these models, an agent is said to know a fact $\varphi$ if $\varphi$ is true in all the worlds she considers possible. While reasoning about knowledge with this semantics has proved useful, as is well known, it suffers from what is known in the literature as the *logical omniscience* problem: under possible-world semantics, agents know all tautologies and know the logical consequences of their knowledge.

While logical omniscience is certainly not always an issue, in many applications it is. For example, in the context of distributed computing, we are interested in polynomial-time algorithms, although in some cases the knowledge needed to perform optimally may require calculations that cannot be performed in polynomial time (unless $P = NP$) [26]; in the context of security, we may want to reason about computationally bounded adversaries who cannot factor a large composite number, and thus cannot be logically omniscient; in game theory, we may be interested in the impact of computational resources on solution concepts (for example, what will agents do if computing a Nash equilibrium is difficult).

Not surprisingly, many approaches for dealing with the logical omniscience problem have been suggested (see [10, Chapter 9] and [25]). A far from exhaustive list of approaches includes:

- syntactic approaches [5,24,18], where an agent's knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);
- *awareness* [7], where an agent knows $\varphi$ if she is aware of $\varphi$ and $\varphi$ is true in all the worlds she considers possible;

☆ A preliminary version of this paper appeared in the *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge*, 2007, pp. 169–176.

* Corresponding author.
*E-mail addresses:* halpern@cs.cornell.edu (J.Y. Halpern), riccardo@ccs.neu.edu (R. Pucella).

0004-3702/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.artint.2010.04.009

- *algorithmic knowledge* [12] where, roughly speaking, an agent knows $\varphi$ if her knowledge algorithm returns "Yes" on a query of $\varphi$; and
- *impossible worlds* [29], where the agent may consider possible worlds that are logically inconsistent (for example, where $p$ and $\neg p$ may both be true).

Which approach is best to use, of course, depends on the application. One goal of this paper is to elucidate the aspects of the application that make a logic more or less appropriate. We start by considering the expressive power of these approaches. In Section 3, we examine the expressiveness of the approaches for a general epistemic logic. It may seem that there is not much to say with regard to expressiveness, since it has been shown that all these approaches are equi-expressive and, indeed, can capture all epistemic states (see [31,10] and Section 2). However, this result holds only if we allow an agent to consider no worlds possible. As we show, this equivalence no longer holds in contexts where agents must consider some worlds possible.

This difference in expressive power is particularly relevant once we have probability in the picture. In Section 4, we examine the logical omniscience problem in the context of an epistemic logic that can talk explicitly about probability, with formulas of the form $K(\ell(\mathit{Prime}_n) = 1/3)$, read "the agent knows that the probability that $\mathit{Prime}_n$ is true is 1/3". We show that in the presence of probabilities, the approaches to dealing with logical omniscience that make sense in this setting are not equi-expressive.

But expressive power is only part of the story. We consider here (mainly by example) the *pragmatics* of dealing with logical omniscience—an issue that has largely been ignored: how to choose an approach and construct an appropriate model. In Section 5, we examine the four main approaches to logical omniscience, and identify some guiding principles for choosing an approach to model a situation, based on the source of the lack of logical omniscience in that situation. Coming up with an appropriate structure can be nontrivial. As a specific contribution, we illustrate a general approach to deriving an impossible-worlds structure based on an implicit description of the situation, which seems to be appropriate for a number of situations of interest.

## 2. The four approaches: a review

We now review the standard possible-worlds approach and the four approaches to dealing with logical omniscience discussed in the introduction. For ease of exposition we focus on the single-agent propositional case. While in many applications it is important to consider more than one agent and to allow first-order features (indeed, this is true in some of our examples), the issues that arise in dealing with multiple agents and first-order features are largely orthogonal to those involved in dealing with logical omniscience. Thus, we do not discuss these extensions here.

### *2.1. The standard approach*

We define a propositional language $\mathcal{L}^K$ of knowledge. Starting with a set $\Phi$ of primitive propositions, we close off under conjunction ($\wedge$), negation ($\neg$), and the $K$ operator. As usual, we consider $\varphi \vee \psi$ to be an abbreviation for $\neg(\neg\varphi \wedge \neg\psi)$, and $\varphi \Rightarrow \psi$ to be an abbreviation for $\neg\varphi \vee \psi$. $K\varphi$ will usually be read as "the agent knows $\varphi$", but because $K\varphi \Rightarrow \varphi$ will not always hold in our models, $K\varphi$ will sometimes have a more natural reading as "the agent believes $\varphi$". None of our results depend on the reading of the operator.

We give semantics to $\mathcal{L}^K$ formulas using Kripke structures. For simplicity, we focus on approaches that satisfy the K45 axioms (as well as KD45 and S5).[1] In this case, a *K45 Kripke structure* is a triple $(W, W', \pi)$, where $W$ is a nonempty set of *possible worlds* (or *worlds*, for short), $W' \subseteq W$ is the set of worlds that the agent considers possible, and $\pi$ is an *interpretation* that associates with each world a truth assignment $\pi(w)$ to the primitive propositions in $\Phi$. Note that the agent need not consider every possible world (that is, each world in $W$) possible. Then we have

$(M, w) \models p$ iff $\pi(w)(p) = \mathbf{true}$, where $p \in \Phi$.
$(M, w) \models \neg\varphi$ iff $(M, w) \not\models \varphi$.
$(M, w) \models \varphi \wedge \psi$ iff $(M, w) \models \varphi$ and $(M, w) \models \psi$.
$(M, w) \models K\varphi$ iff $(M, w') \models \varphi$ for all $w' \in W'$.

This semantics suffers from the logical omniscience problem. In particular, one sound axiom is

$$\big(K\varphi \wedge K(\varphi \Rightarrow \psi)\big) \Rightarrow K\psi,$$

which says that an agent's knowledge is closed under implication. In addition, the *knowledge generalization* inference rule is sound:

From $\varphi$ infer $K\varphi$.

[1] We could extend the investigation in this paper to more general structures satisfying weaker axioms, but consider the more standard setting suffices for the points we want to make. We expect similar results to the ones we obtain here to hold for more general structures, but have not checked the details.

Thus, agents know all tautologies. As is well known, two other axioms are sound in K45 Kripke structures:

$$K\varphi \Rightarrow KK\varphi$$

and

$$\neg K\varphi \Rightarrow K\neg K\varphi.$$

These are known respectively as the positive and negative introspection axioms. (These properties characterize K45.)

In the structures we consider, we allow $W'$ to be empty, in which case the agent does not consider any worlds possible. In such structures, $K\varphi$ is true for all $\varphi$, including *false*. A *KD45 Kripke structure* is a K45 Kripke structure $(W, W', \pi)$ where $W' \neq \emptyset$. Thus, in a KD45 Kripke structure, the agent always considers at least one world possible. In KD45 Kripke structures, the axiom

$$\neg K(\mathit{false})$$

is sound, which implies that the agent cannot know inconsistent facts. The logic KD45 results when we add this axiom to K45. *S5 Kripke structures* are KD45 Kripke structures where $W = W'$; that is, the agent considers all worlds in $W$ possible. In S5 Kripke structures, the axiom

$$K\varphi \Rightarrow \varphi,$$

which says that the agent can know only true facts, is sound. Adding this axiom to the KD45 axioms gives us the logic S5.

### 2.2. The syntactic approach

The intuition behind the syntactic approach for dealing with logical omniscience is simply to explicitly list, at every possible world $w$, the set of formulas that the agent knows at $w$. A *syntactic structure* has the form $M = (W, W', \pi, \mathcal{C})$, where $(W, W', \pi)$ is a K45 Kripke structure and $\mathcal{C}$ associates a set of formulas $\mathcal{C}(w)$ with every world $w \in W$. The semantics of primitive propositions, conjunction, and negation is just the same as for Kripke structures. For knowledge, we have

$$(M, w) \models K\varphi \quad \text{iff} \quad \varphi \in \mathcal{C}(w).$$

Of course, for syntactic structures, the set of possible worlds plays no role in the semantics of knowledge.

### 2.3. Awareness

Awareness is based on the intuition that an agent should be aware of a concept before she can know it. The formulas that an agent is aware of are represented syntactically; we associate with every world $w$ the set $\mathcal{A}(w)$ of formulas that the agent is aware of. For an agent to know a formula $\varphi$, not only does $\varphi$ have to be true at all the worlds she considers possible, but she has to be aware of $\varphi$ as well. A *K45 awareness structure* is a tuple $M = (W, W', \pi, \mathcal{A})$, where $(W, W', \pi)$ is a K45 Kripke structure and $\mathcal{A}$ maps worlds to sets of formulas. We now define

$$(M, w) \models K\varphi \quad \text{iff} \quad (M, w') \models \varphi \text{ for all } w' \in W' \text{ and } \varphi \in \mathcal{A}(w).$$[2]

We can define KD45 and S5 awareness structures in the obvious way: $M = (W, W', \pi, \mathcal{A})$ is a KD45 awareness structure when $(W, W', \pi)$ is a KD45 structure, and an S5 awareness structure when $(W, W', \pi)$ is an S5 structure.

### 2.4. Algorithmic knowledge

In some applications, there is a computational intuition underlying what an agent knows; that is, an agent computes what she knows using an algorithm. *Algorithmic knowledge* is one way of formalizing this intuition. An *algorithmic knowledge structure* is a tuple $M = (W, W', \pi, \mathtt{A})$, where $(W, W', \pi)$ is a K45 Kripke structure and $\mathtt{A}$ is a *knowledge algorithm* that returns "Yes", "No", or "?" given a formula $\varphi$.[3] Intuitively, $\mathtt{A}(\varphi)$ returns "Yes" if the agent can compute that $\varphi$ is true, "No" if the agent can compute that $\varphi$ is false, and "?" otherwise. In algorithmic knowledge structures,

$$(M, w) \models K\varphi \quad \text{iff} \quad \mathtt{A}(\varphi) = \text{“Yes”}.$$

[2] In [7], the symbol $K$ is reserved for the standard definition of knowledge; the definition we have just given is denoted as $X\varphi$, where $X$ stands for *explicit* knowledge. A similar remark applies to the algorithmic knowledge approach below. We use $K$ throughout for ease of exposition.

[3] In [12], the knowledge algorithm is also given an argument that describes the agent's local state, which, roughly speaking, captures the relevant information that the agent has. However, in our single-agent static setting, there is only one local state, so this argument is unneeded.

As in the syntactic approach, the set of possible worlds plays no role in the semantics of knowledge.

An important class of knowledge algorithms consists of the *sound* knowledge algorithms. When a sound knowledge algorithm returns "Yes" to a query $\varphi$, then the agent knows (in the standard sense) $\varphi$, and when it returns "No" to a query $\varphi$, then the agent does not know (again, in the standard sense) $\varphi$. Thus, if $\mathtt{A}$ is a sound knowledge algorithm, then $\mathtt{A}(\varphi) =$ "Yes" implies $(M, w) \models \varphi$ for all $w \in W'$, and $\mathtt{A}(\varphi) =$ "No" implies there exists $w \in W'$ such that $(M, w) \models \neg\varphi$. (When $\mathtt{A}(\varphi) =$ "?", nothing is prescribed.)

### *2.5. Impossible worlds*

The impossible-worlds approach relies on relaxing the notion of possible world. Take the special case of logical omniscience that says that an agent knows all tautologies. This is a consequence of the fact that a tautology must be true at every possible world. Thus, one way to eliminate this problem is to allow tautologies to be false at some worlds. Clearly, those worlds do not obey the usual laws of logic—they are *impossible possible worlds* (or *impossible worlds*, for short).

A *K45* (resp., *KD45*, *S5*) *impossible-worlds structure* is a tuple $M = (W, W', \pi, \mathcal{C})$, where $(W, W' \cap W, \pi)$ is a K45 (resp., KD45, S5) Kripke structure, $W'$ is the set of worlds that the agent considers possible, and $\mathcal{C}$ associates with each world in $W' - W$ a set of formulas. $W'$, the set of worlds the agent considers possible, is not required to be a subset of $W$—the agent may well include impossible worlds in $W'$. The worlds in $W' - W$ are the impossible worlds. We can also consider a class of impossible-worlds structures intermediate between K45 and KD45 impossible-worlds structures. A *KD45⁻ impossible-worlds structure* is a K45 impossible-worlds structure $(W, W', \pi, \mathcal{C})$ where $W'$ is nonempty. In a KD45$^-$ impossible-worlds structure, we do not require that $W' \cap W$ be nonempty.

A formula $\varphi$ is true at a world $w \in W' - W$ if and only if $\varphi \in \mathcal{C}(w)$; for worlds $w \in W$, the truth assignment is like that in Kripke structures. Thus,

- if $w \in W$, then $(M, w) \models p$ iff $\pi(w)(p) = \mathbf{true}$;
- if $w \in W$, then $(M, w) \models K\varphi$ iff $(M, w') \models \varphi$ for all $w' \in W'$;
- if $w \in W' - W$, then $(M, w) \models \varphi$ iff $\varphi \in \mathcal{C}(w)$.

We remark that when we speak of validity in impossible-worlds structures, we mean truth at all possible worlds in $W$ in all impossible-worlds structures $M = (W, \ldots)$.

## 3. Expressive power

There is a sense in which all four approaches are equi-expressive, and can capture all states of knowledge. To make this precise, define a set $\Phi'$ of formulas in $\mathcal{L}^K$ to be *propositionally consistent* if $\Phi'$ is a consistent set of formulas of propositional logic when we treat formulas of the form $K\varphi$ as primitive propositions (a distinct one for each $\varphi$). Thus, propositional consistency ignores properties of the knowledge operator. We take for granted here a sound and complete axiomatization of propositional logic [6], and therefore what we call a propositionally consistent set is also a propositionally satisfiable set.

**Theorem 3.1.** *(See [31,10].) For every finite set $F$ of formulas and every propositionally consistent set $G$ of formulas, there exists a syntactic structure (resp., K45 awareness structure, KD45$^-$ impossible-worlds structure, algorithmic knowledge structure) $M = (W, \ldots)$ and a world $w \in W$ such that $(M, w) \models K\varphi$ if and only if $\varphi \in F$, and $(M, w) \models \psi$ for all $\psi \in G$.*[4]

**Proof.** We review the basic idea of the proof, since it will set the stage for our later results.

- For syntactic structures, let $M = (\{w\}, \varnothing, \pi, \mathcal{C})$, where $\mathcal{C}(w) = F$ and $\pi(w)$ is such that $(M, w) \models \psi$ for all $\psi \in G$. (Since $G$ is propositionally consistent, there must be a truth assignment that makes all the formulas in $G$ true; we can take $\pi(w)$ to be that truth assignment.)
- For K45 awareness structure, let $M = (\{w\}, \varnothing, \pi, \mathcal{A})$, where $\mathcal{A}(w) = F$ and $\pi(w)$ makes all the formulas in $G$ true.
- For KD45$^-$ impossible-worlds structure, let $M = (\{w\}, \{w'\}, \pi, \mathcal{C})$, where $\mathcal{C}(w') = F$ and $\pi(w)$ makes all the formulas in $G$ true.
- For algorithmic knowledge, let $M = (\{w\}, \varnothing, \pi, \mathtt{A})$, where $\mathtt{A}(\varphi) =$ "Yes" iff $\varphi \in F$ and $\pi(w)$ makes all the formulas in $G$ true. $\square$

Despite the name, the introspective axioms of K45 are not valid in K45 awareness structures or K45 impossible-worlds structures. Indeed, it follows from Theorem 3.1 that no axioms of knowledge are valid in these structures. (Take $F$ to be the empty set.) In fact, we can show that a formula is valid with respect to K45 awareness structures (or KD45 or KD45$^-$ impossible-worlds structures, syntactic structures, algorithmic knowledge structures) if and only if it is a substitution

[4] This result extends to infinite sets $F$ of formulas for syntactic structures, K45 awareness structures, and KD45$^-$ impossible-worlds structures. For algorithmic knowledge structures, the result extends to recursive sets $F$ of formulas.

instance of a propositional tautology, that is, the result of substituting arbitrary formulas in $\mathcal{L}^K$ for the primitive propositions in a propositional tautology. To make this precise, let *Prop* be the axiom

$$\varphi \text{ is a substitution instance of a valid formula of propositional logic} \qquad (Prop)$$

and *MP* be the inference rule

$$\text{From } \varphi \Rightarrow \psi \text{ and } \varphi \text{ infer } \psi. \qquad (MP)$$

The inference rule *MP* is not needed for our next result, but it will be needed later when we look at axiomatizations for $\mathcal{L}^K$ in different structures, so we introduce it here.

**Theorem 3.2.** $\{Prop, MP\}$ *is a sound and complete axiomatization of* $\mathcal{L}^K$ *with respect to K45 awareness structures (resp., K45 and KD45*$^-$ *impossible-worlds structures, syntactic structures, algorithmic knowledge structures).*

**Proof.** Soundness is straightforward to establish in all cases. For completeness, we show that a consistent formula is satisfiable. Suppose that $\varphi$ is consistent with $\{Prop, MP\}$. It suffices to show that $\varphi$ is satisfiable in a K45 awareness (resp., K45 and KD45$^-$ impossible-worlds structure, syntactic structure, algorithmic knowledge structure). Viewing formulas of the form $K\psi$ as primitive propositions, $\varphi$ must be propositionally consistent. Thus, there must be a truth assignment $v$ to the primitive propositions and formulas of the form $K\psi$ that appear in $\varphi$ such that $\varphi$ evaluates to true under this truth assignment. Let $F$ consist of all formulas $\psi$ such that $v(K\psi) = \textbf{true}$ and let $G$ consist of all the propositional formulas $\psi$ such that $v(\psi) = \textbf{true}$. Let $M$ be the structure guaranteed to exist by Theorem 3.1. It is easy to see that $(M, w) \models \varphi$. □

It follows from Theorem 3.2 that a formula is valid with respect to K45 awareness structures (resp., K45 and KD45$^-$ impossible-worlds structures, syntactic structures, algorithmic knowledge structures) if and only if it is a substitution instance of a propositional tautology. Thus, deciding if a formula is valid is co-NP complete, just as it is for propositional logic.

Theorems 3.1 and 3.2 rely on the fact that we are considering K45 awareness structures and KD45$^-$ (or K45) impossible-worlds structures. (Whether we consider K45, KD45, or S5 is irrelevant in the case of syntactic structures and algorithmic knowledge structures, since the truth of a formula does not depend on what worlds an agent considers possible.) There are constraints on what can be known if we consider KD45 and S5 awareness structures and impossible-worlds structures. The constraints depend on which structures we consider. To make the constraints precise, we need a few definitions. We say a set of formulas $F$ is *downward closed* if the following conditions hold:

(a) if $\varphi \wedge \psi \in F$, then both $\varphi$ and $\psi$ are in $F$;
(b) if $\neg\neg\varphi \in F$, then $\varphi \in F$;
(c) if $\neg(\varphi \wedge \psi) \in F$, then either $\neg\varphi \in F$ or $\neg\psi \in F$ (or both); and
(d) if $K\varphi \in F$, then $\varphi \in F$.

We say that $F$ is *k-compatible* with $F'$ if $K\psi \in F'$ implies that $\psi \in F$.

**Proposition 3.3.** *Suppose that* $M = (W, W', \ldots)$ *is a KD45 awareness structure,* $w \in W$*, and* $w' \in W'$*. Let* $F = \{\varphi \mid (M, w) \models K\varphi\}$ *and let* $F' = \{\psi \mid (M, w') \models \psi\}$*. Then* $F'$ *is a propositionally consistent downward-closed set of formulas that contains* $F$*.*

**Proof.** Suppose that $M = (W, W', \ldots)$ is a KD45 awareness structure. Let $w$, $w'$, $F$, and $F'$ be as in the statement of the theorem. Clearly $F \subseteq F'$. Since $w'$ is a possible world, it is easy to see that $F'$ satisfies the first three conditions of being downward closed. For the last condition, note that if $(M, w') \models K\psi$, then we must have $(M, w'') \models \psi$ for all worlds $w'' \in W'$, so $(M, w') \models \psi$. Finally, $F'$ must be propositionally consistent, since $w'$ is a possible world. □

**Proposition 3.4.** *Suppose that* $M = (W, W', \ldots)$ *is a KD45 impossible-worlds structure,* $w \in W$*, and* $w' \in W \cap W'$*. Let* $F = \{\varphi \mid (M, w) \models K\varphi\}$ *and let* $F' = \{\psi \mid (M, w') \models \psi\}$*. Then*

(a) $F'$ *is a propositionally consistent downward-closed set of formulas that contains* $F$;
(b) $F$ *is k-compatible with* $F'$.

**Proof.** The argument for (a) is the same as in the proof of Proposition 3.3, since $w' \in W \cap W'$ in this case. For (b), to see that $F$ is k-compatible with $F'$, suppose that $K\varphi \in F'$. By the definition of $F'$, this means that $(M, w') \models K\varphi$. It follows that $(M, w'') \models \varphi$ for all $\varphi \in W'$. Hence, $(M, w) \models K\varphi$, so $\varphi \in F$. (Note that this argument does not work for awareness structures, since we may not have $\varphi \in \mathcal{A}(w)$, and therefore we cannot necessarily derive $(M, w) \models K\varphi$.) □

The next result shows that the constraints on $F$ described in Propositions 3.3 and 3.4 are the only constraints on $F$.

**Theorem 3.5.** *If $F$ and $F'$ are such that $F'$ is propositionally consistent downward-closed set of formulas that contains $F$, then there exists a KD45 awareness structure $M = (\{w, w'\}, \{w'\}, \pi, \mathcal{A})$ such that $(M, w) \models K\varphi$ iff $\varphi \in F$ and $(M, w') \models \psi$ for all $\psi \in F'$. If, in addition, $F$ is k-compatible with $F'$, then there exists a KD45 impossible-worlds structure $M = (\{w, w'\}, \{w', w''\}, \pi, \mathcal{C})$ such that $(M, w) \models K\varphi$ iff $\varphi \in F$ and $(M, w') \models \psi$ for all $\psi \in F'$. Finally, if $F = F'$, then we can take $w = w'$, so that $M$ is an S5 awareness (resp., S5 impossible-worlds) structure.*

**Proof.** In the case of KD45 awareness structures, let $M = (\{w, w'\}, \{w'\}, \pi, \mathcal{A})$, where $\pi(w')$ makes all the propositional formulas in $F'$ true, $\mathcal{A}(w) = F$, and $\mathcal{A}(w') = \{\varphi \mid K\varphi \in F'\}$. We now prove by induction that if $\varphi \in F'$ then $(M, w') \models \varphi$. This is true by construction in the case of primitive propositions and follows easily from the induction hypothesis in the case of conjunctions. If $\varphi$ has the form $K\psi$ then, since $\psi$ must be in $F'$, it follows from the induction hypothesis that $(M, w') \models \psi$ and, by construction, that $\psi \in \mathcal{A}(w')$. Thus, $(M, w') \models K\psi$. Finally, if $\varphi$ has the form $\neg\psi$, we consider the possible forms of $\psi$. If $\psi$ is a primitive proposition it follows from the definition of $\pi(w')$. If $\psi$ has the form $\neg\psi'$, then $\psi' \in F'$, so, by the induction hypothesis, $(M, w') \models \psi'$. Hence, $(M, w') \models \varphi$. Similarly, the result follows from the definition of downward closure and the induction hypothesis if $\psi$ has the form $\psi_1 \wedge \psi_2$. Finally, if $\psi$ has the form $K\psi'$, then the result follows from the definition on $\mathcal{A}(w')$. It is now immediate that $(M, w) \models K\varphi$ iff $\varphi \in F$: if $(M, w) \models K\varphi$ then it follows from the definition of $\mathcal{A}(w)$ that we must have $\varphi \in F$. Conversely, if $\varphi \in F$, then $\varphi \in \mathcal{A}(w)$ and $(M, w') \models \varphi$ (since $F \subseteq F'$), so $(M, w) \models K\varphi$.

If $F = F'$, then we can take $w = w'$ in this argument to get an S5 awareness structure.

In the case of impossible-worlds structures, let $M = (\{w, w'\}, \{w', w''\}, \pi, \mathcal{C})$, where $\pi(w')$ makes all the propositional formulas in $F'$ true and $\mathcal{C}(w'') = F$. A proof by induction on the structure of formulas much like that above shows that $(M, w') \models \varphi$ if $\varphi \in F'$. To deal with the case that $\varphi = K\psi$, we use the fact that $F$ is k-compatible with $F'$ to get that $\psi \in F$, so that $(M, w'') \models \psi$. To see that $(M, w) \models K\varphi$ iff $\varphi \in F$, first observe that if $\varphi \in F$ then, by construction $\varphi \in \mathcal{C}(w'')$, and, since $F \subseteq F'$, $(M, w') \models \varphi$, so $(M, w) \models K\varphi$. For the converse, if $(M, w) \models K\varphi$, then $(M, w'') \models \varphi$, so $\varphi \in F$. □

We can characterize these properties axiomatically. Let (*Ver*) be the standard *Veridicality* axiom, which says that everything known must be true:

$$K\varphi \Rightarrow \varphi. \tag{Ver}$$

Let $\mathrm{AX}_{Ver}$ be the axiom system consisting of {*Prop*, *MP*, *Ver*}. The fact that the set of formulas known must be a subset of a downward-closed set is characterized by the following axiom:

$$\neg(K\varphi_1 \wedge \cdots \wedge K\varphi_n) \quad \text{if } \mathrm{AX}_{Ver} \vdash \neg(\varphi_1 \wedge \cdots \wedge \varphi_n). \tag{DC}$$

The key point here is that, as we shall show, a propositionally consistent set of formulas that is downward closed must be consistent with $\mathrm{AX}_{Ver}$.

The fact that the set of formulas that is known is k-compatible with a downward-closed set of formulas is characterized by the following axiom:

$$(K\varphi_1 \wedge \cdots \wedge K\varphi_n) \Rightarrow (K\psi_1 \vee \cdots \vee K\psi_m) \quad \text{if } \mathrm{AX}_{Ver} \vdash \varphi_1 \wedge \cdots \wedge \varphi_n \Rightarrow (K\psi_1 \vee \cdots \vee K\psi_m). \tag{KC}$$

Axiom *DC* is just the special case of axiom *KC* where $m = 0$. It is also easy to see that *KC* (and therefore *DC*) follow from *Ver*. As we now show, *DC* is strictly weaker than *KC*. Suppose that $\Phi = \{p\}$ (so that $p$ is the only primitive proposition in the language), and take $M = (\{w_1, w_2\}, w_2, \pi, \mathcal{A})$, where $\pi$ makes $p$ true at both $w_1$ and $w_2$, $\mathcal{A}(w_1) = \{Kp\}$, and $\mathcal{A}(w_2) = \{Kp, p\}$. Every instance of *DC* is valid in $M$ (this is the soundness part of Theorem 3.6(a) below). Now consider the formula $Kp \Rightarrow Kp$. This is an instance of *Prop*, so $\mathrm{AX}_{Ver} \vdash Kp \Rightarrow Kp$. Thus, if every instance of *KC* were valid in $M$, we would have $KKp \Rightarrow Kp$ valid in $M$. It is easy to check that $(M, w_1) \models KKp$, since $Kp \in \mathcal{A}(w_1)$ and $(M, w_2) \models Kp$, but $(M, w_1) \models \neg Kp$, because $p \notin \mathcal{A}(w_1)$. Therefore the $KKp \Rightarrow Kp$ instance of *KC* is not in fact valid in $M$.

Let $\mathrm{AX}_{DC} = \{Prop, MP, DC\}$ and let $\mathrm{AX}_{KC} = \{Prop, MP, KC\}$.

**Theorem 3.6.**

(a) *$AX_{DC}$ is a sound and complete axiomatization of $\mathcal{L}^K$ with respect to KD45 awareness structures*;
(b) *$AX_{KC}$ is a sound and complete axiomatization of $\mathcal{L}^K$ with respect to KD45 impossible-worlds structures*;
(c) *$AX_{Ver}$ is a sound and complete axiomatization of $\mathcal{L}^K$ with respect to S5 awareness structures and S5 impossible-worlds structures.*

**Proof.** (a) We first prove soundness. Consider axiom *DC*. Suppose that $\mathrm{AX}_{Ver} \vdash \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$. Let $M = (W, W', \pi, \mathcal{A})$ be a KD45 awareness structure. For each world $w' \in W'$, it easily follows from Proposition 3.3 (taking $w = w'$) that each instance of axiom *Ver* holds at $(M, w')$, as does each instance of *Prop*. An easy argument by induction on the length of proof then shows that, if $\mathrm{AX}_{Ver} \vdash \psi$, then $(M, w') \models \psi$. In particular, $(M, w') \models \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$. It follows that, for each $w \in W$, we must have $(M, w) \models \neg(K\varphi_1 \wedge \cdots \wedge K\varphi_n)$.

For completeness, it suffices to show that, given an $\mathrm{AX}_{DC}$-consistent formula $\varphi$, there exists a KD45 awareness structure $M$ and world $w$ such that $(M, w) \models \varphi$. So suppose that $\varphi$ is $\mathrm{AX}_{DC}$-consistent. Let $G$ be a maximal $\mathrm{AX}_{DC}$-consistent set containing $\varphi$. Let $F = \{\psi \mid K\psi \in G\}$. We claim that $F$ is $AX_{Ver}$-consistent. If not, then there exist $\varphi_1, \ldots, \varphi_n \in G$ such that $AX_{Ver} \vdash \neg(\varphi_1 \wedge \cdots \wedge \varphi_n)$. But then by axiom *DC*, we have that $AX_{DC} \vdash \neg(K\varphi_1 \wedge \cdots \wedge K\varphi_n)$, contradicting the fact that $G$ is $AX_{DC}$-consistent. Thus, $F$ is consistent with $AX_{Ver}$. Let $F'$ be a maximal $AX_{Ver}$-consistent set extending $F$. Then it is easy to check that $F'$ is a propositionally consistent downward-closed set of formulas that contains $F$. Thus, by Theorem 3.5, there is a KD45 awareness structure $M = (\{w, w'\}, \{w'\}, \pi, \mathcal{A})$ such that $(M, w) \models K\psi$ for all $\psi \in F$. We can assume without loss of generality that $w \neq w'$ and that $\pi(w)$ makes all the primitive propositions in $F$ true. (Note that this would not be the case if we were dealing with S5 awareness structures.) An easy induction on the structure of formulas then shows that $(M, w) \models \psi$ for all $\psi \in G$. In particular, $(M, w) \models \varphi$.

(b) For soundness, essentially the same argument as in part (a) shows that axiom *DC* is sound in KD45 impossible-worlds structures. A similar argument also shows the soundness of *KC* with respect to KD45 impossible-worlds structures. For suppose that $M = (W, W', \pi, \mathcal{C})$ is an impossible-worlds structure, $w \in W$, $AX_{Ver} \vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \Rightarrow (K\psi_1 \vee \cdots \vee K\psi_m)$, and $(M, w) \models K\varphi_1 \wedge \cdots \wedge K\varphi_n$. Thus, $(M, w'') \models \varphi_1 \wedge \cdots \wedge \varphi_n$ for all $w'' \in W'$. But since each world in $W \cap W'$ is a model of $AX_{Ver}$, if $w' \in W \cap W'$, we must have $(M, w') \models K\psi_1 \vee \cdots \vee K\psi_m$. Moreover, since $W \cap W' \neq \varnothing$, there must be some world $w' \in W \cap W'$. It follows that, for some $j \in \{1, \ldots, m\}$, $(M, w') \models K\psi_j$. Thus, $(M, w'') \models \psi_j$ for all $w'' \in W'$, so $(M, w) \models K\psi_j$. It follows that $(M, w) \models K\psi_1 \vee \cdots \vee K\psi_m$, as desired.

For completeness, we use much the same argument as in part (a). Suppose that $\varphi$ is $AX_{KC}$-consistent. Let $G$ be a maximal $AX_{KC}$-consistent set containing $\varphi$. Let $F = \{\psi \mid K\psi \in G\}$, and let $G' = F \cup \{\neg\psi \mid \neg K\psi \in G\}$. We again claim that $G'$ is $AX_{Ver}$-consistent. If not, then there exist $K\varphi_1, \ldots, K\varphi_n, K\psi_1, \ldots, K\psi_m \in G$ such that $AX_{Ver} \vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \Rightarrow (K\psi_1 \vee \cdots \vee K\psi_m)$. By axiom *KC*, we have that $AX_{KC} \vdash (K\varphi_1 \wedge \cdots \wedge K\varphi_n) \Rightarrow (K\psi_1 \vee \cdots \vee K\psi_m)$, contradicting the fact that $G$ is $AX_{KC}$-consistent. Thus, $G'$ is consistent with $AX_{Ver}$. Again, let $F'$ be a maximal $AX_{Ver}$-consistent set extending $G'$. Then it is easy to check that $F'$ is a propositionally consistent downward-closed set of formulas that contains $F$; moreover the construction guarantees that $F$ is k-compatible with $F'$. Thus, by Theorem 3.5, there is a KD45 impossible-worlds structure $M = (\{w, w'\}, \{w', w''\}, \pi, \mathcal{C})$ such that $(M, w) \models K\psi$ for all $\psi \in F$. We can assume without loss of generality that $w \neq w'$ and that $\pi(w)$ makes all the primitive propositions in $F$ true. An easy induction on the structure of formulas then shows that $(M, w) \models \psi$ for all $\psi \in G$. In particular, $(M, w) \models \varphi$.

(c) For soundness, as we have already observed, the soundness of *Ver* in S5 awareness and impossible-worlds structures follows easily from Propositions 3.3 and 3.4.

For completeness, let $AX = \{Prop, MP, Ver\}$. Suppose that $\varphi$ is consistent with AX. Extend $\varphi$ to a maximally AX-consistent set $F$ of formulas. It suffices to show that $F$ is satisfiable in an S5 awareness structure and in an S5 impossible-worlds structure. In the case of awareness structures, consider the structure $M = (\{w\}, \{w\}, \pi, \mathcal{A})$, where $\pi(w)(p) = \mathbf{true}$ iff $p \in F$ and $\mathcal{A}(w) = \{\psi \mid K\psi \in F\}$. We now show by induction on the structure of formulas that $(M, w) \models \psi$ iff $\psi \in F$. If $\psi$ is a primitive proposition, then this is immediate from the definition of $\pi$. If $\psi$ has the form $\neg\psi'$, then the result is immediate from the induction hypothesis. If $\psi$ has the form $\psi_1 \wedge \psi_2$, this is immediate from the observation that, since $F$ is a maximal AX-consistent set and propositional reasoning is sound in AX that $\psi_1 \wedge \psi_2 \in F$ iff $\psi_1 \in F$ and $\psi_2 \in F$. If $\psi$ has the form $K\psi'$, note that if $K\psi' \in F$ then $\psi' \in F$ (since $Ver \in F$). By the induction hypothesis, $(M, w) \models \psi'$. Thus, $(M, w) \models K\psi'$. For the converse, if $(M, w) \models K\psi'$, suppose, by way of contradiction, that $K\psi' \notin F$. Then, by construction, $\psi' \notin \mathcal{A}(w)$. Thus, $(M, w) \models \neg K\psi'$, a contradiction.

To show that $F$ is satisfiable in an S5 impossible-worlds structure, consider the structure $M = (\{w\}, \{w, w'\}, \pi, \mathcal{C})$, where $\pi(w)(p) = \mathbf{true}$ iff $p \in F$ and $\mathcal{C}(w') = \{\psi \mid K\psi \in F\}$. Thus, $\mathcal{C}(w')$ is the same set of formulas as $\mathcal{A}(w)$ in the argument for S5 awareness structures. An almost identical argument as in the case of S5 awareness structures now shows that $(M, w) \models \psi$ iff $\psi \in F$. We leave details to the reader. □

Note that in all cases of Theorem 3.6, we proved completeness by constructing, for a given consistent formula, a satisfying structure with few worlds. This indicates that awareness structures and impossible-worlds structures are quite flexible—the lack of restrictions on both awareness sets and truth assignments to impossible worlds lets us easily capture states of knowledge with few worlds.

**Corollary 3.7.** *The satisfiability problem for the language $\mathcal{L}^K$ with respect to KD45 awareness structures (resp., KD45 impossible-worlds structures, S5 awareness structures, S5 impossible-worlds structures) is NP-complete.*

**Proof.** NP-hardness follows immediately from the observation that $\mathcal{L}^K$ contains propositional logic. The fact that the satisfiability problem with respect to each of these classes of structures is in NP follows from the construction of Theorem 3.6, which shows that if a formula $\varphi$ is satisfiable with respect to KD45 awareness structures (resp., KD45 impossible-worlds structures, S5 awareness structures, S5 impossible-worlds structures), then it is consistent with respect to $AX_{DC}$ (resp. $AX_{KC}$, $AX_{Ver}$), which in turn implies that it is satisfiable in a KD45 awareness structure (resp., KD45 impossible-worlds structure, S5 awareness structure, S5 impossible-worlds structure) $M = (W, W', \ldots)$ with two (resp., three, one) world(s). Without loss of generality, we can also assume that, in the case of awareness structures, at each world $w \in W$, $\mathcal{A}(w)$ is a subset of $Sub(\varphi)$, the set of subformulas of $\varphi$, and $\pi(w)(p) = \mathbf{true}$ only if $p$ is a subformula of $\varphi$; similarly, in the case of impossible-worlds structures, we can assume that for each impossible world $w'$, $\mathcal{C}(w')$ is a subset of the subformulas of $\varphi$. (If this is not true in $M$, then we can easily modify $M$ so that this is true without affecting the truth of $\varphi$ or any subformula of $\varphi$ in any world.) Thus, we can guess a satisfying structure for $\varphi$ and verify that it satisfies $\varphi$ in time linear in the length of $\varphi$. □

## 4. Adding probability

While the differences between K45, KD45$^-$, and KD45 impossible-worlds structures may appear minor, they turn out to be important when we add probability to the picture. As pointed out by Cozic [2], standard models for reasoning about probability suffer from the same logical omniscience problem as models for knowledge. In the language considered by Fagin, Halpern, and Megiddo [9] (FHM from now on), there are formulas that talk explicitly about probability. A formula such as $\ell(Prime_n) = 1/3$ says that the probability that $n$ is prime is 1/3. In the FHM semantics, a probability is put on the set of worlds that the agent considers possible. The probability of a formula $\varphi$ is then the probability of the set of worlds where $\varphi$ is true. Clearly, if $\varphi$ and $\psi$ are logically equivalent, then $\ell(\varphi) = \ell(\psi)$ will be true. However, the agent may not recognize that $\varphi$ and $\psi$ are equivalent, and so may not recognize that $\ell(\varphi) = \ell(\psi)$. Problems of logical omniscience with probability can to some extent be reduced to problems of logical omniscience with knowledge in a logic that combines knowledge and probability [8]. For example, the fact that an agent may not recognize $\ell(\varphi) = \ell(\psi)$ when $\varphi$ and $\psi$ are equivalent just amounts to saying that if $\varphi \Leftrightarrow \psi$ is valid, then we do not necessarily want $K(\ell(\varphi) = \ell(\psi))$ to hold. However, adding knowledge and awareness does not prevent $\ell(\varphi) = \ell(\psi)$ from holding. This is not really a problem if we interpret $\ell(\varphi)$ as the objective probability of $\varphi$; if $\varphi$ and $\psi$ are equivalent, it is an objective fact about the world that their probabilities are equal, so $\ell(\varphi) = \ell(\psi)$ should hold. On the other hand, if $\ell(\varphi)$ represents the agent's subjective view of the probability of $\varphi$, then we do not want to require $\ell(\varphi) = \ell(\psi)$ to hold. This cannot be captured in all approaches.

To make this precise, we first clarify the logic we have in mind. Let $\mathcal{L}^{K,QU}$ be $\mathcal{L}^K$ extended with linear inequality formulas involving probability (called likelihood formulas), in the style of FHM. A likelihood formula is of the form $a_1\ell(\varphi_1) + \cdots + a_n\ell(\varphi_n) \geqslant c$, where $a_1, \ldots, a_n$ and $c$ are integers. (For ease of exposition, we restrict the $\varphi_1, \ldots, \varphi_n$ appearing in likelihood formulas to be propositional, that is, with no occurrences of $\ell$ and $K$; however, the techniques presented here can be extended to deal with formulas that allow arbitrary nesting of $\ell$ and $K$.) We give semantics to these formulas by extending Kripke structures with a probability distribution over the worlds that the agent considers possible. A *probabilistic KD45* (*resp., S5*) *Kripke structure* is a tuple $(W, W', \pi, \mu)$, where $(W, W', \pi)$ is KD45 (resp., S5) Kripke structure, and $\mu$ is a probability distribution over $W'$. To interpret likelihood formulas, we first define $[\![\varphi]\!]_M = \{w \in W \mid \pi(w)(\varphi) = \mathbf{true}\}$, for a propositional formula $\varphi$. We then extend the semantics of $\mathcal{L}^K$ with the following rule for interpreting likelihood formulas:

$$(M, w) \models a_1\ell(\varphi_1) + \cdots + a_n\ell(\varphi_n) \geqslant c \quad \text{iff} \quad a_1\mu\big([\![\varphi_1]\!]_M \cap W'\big) + \cdots + a_n\mu\big([\![\varphi_n]\!]_M \cap W'\big) \geqslant c.$$

Note that the truth of a likelihood formula at a world does not depend on that world; if a likelihood formula is true at a world of a structure $M$, then it is true at every world of $M$.

FHM give an axiomatization for likelihood formulas in probabilistic structures. Aside from propositional reasoning axioms, one axiom captures reasoning with linear inequalities. A *basic inequality formula* is a formula of the form $a_1x_1 + \cdots + a_kx_k + a_{k+1} \leqslant b_1y_1 + \cdots + b_my_m + b_{m+1}$, where $x_1, \ldots, x_k, y_1, \ldots, y_m$ are (not necessarily distinct) variables. A *linear inequality formula* is a Boolean combination of basic linear inequality formulas. A linear inequality formula is valid if the resulting inequality holds under every possible assignment of real numbers to variables. For example, the formula $(2x + 3y \leqslant 5z) \wedge (x - y \leqslant 12z) \Rightarrow (3x + 2y \leqslant 17z)$ is a valid linear inequality formula. To get an instance of *Ineq*, we replace each variable $x_i$ that occurs in a valid formula about linear inequalities by a likelihood term of the form $\ell(\psi)$ (naturally, each occurrence of the variable $x_i$ must be replaced by the same primitive expectation term $\ell(\psi)$). (We can replace *Ineq* by a sound and complete axiomatization for Boolean combinations of linear inequalities; one such axiomatization is given in FHM.)

The other axioms of FHM are specific to probabilistic reasoning, and capture the defining properties of probability distributions:

$$\ell(true) = 1,$$
$$\ell(\neg\varphi) = 1 - \ell(\varphi),$$
$$\ell(\varphi \wedge \psi) + \ell(\varphi \wedge \neg\psi) = \ell(\varphi).$$

It is straightforward to extend all the approaches in Section 2 to the probabilistic setting. In this section, we only consider probabilistic awareness structures and probabilistic impossible-worlds structures, because the interpretation of both algorithmic knowledge and knowledge in syntactic structures does not depend on the set of worlds or any probability distribution over the set of worlds.

A *KD45* (*resp., S5*) *probabilistic awareness structure* is a tuple $(W, W', \pi, \mathcal{A}, \mu)$ where $(W, W', \pi, \mathcal{A})$ is a KD45 (resp., S5) awareness structure and $\mu$ is a probability distribution over the worlds in $W'$. Similarly, a *KD45$^-$* (*resp., KD45, S5*) *probabilistic impossible-worlds structure* is a tuple $(W, W', \pi, \mathcal{C}, \mu)$ where $(W, W', \pi, \mathcal{C})$ is a KD45$^-$ (resp., KD45, S5) impossible-worlds structure and $\mu$ is a probability distribution over the worlds in $W'$. Since the set of worlds that are assigned probability must be nonempty, when dealing with probability, we must restrict to KD45 awareness structures and KD45$^-$ impossible-worlds structures, extended with a probability distribution over the set of worlds the agent considers possible. As we now show, adding probability to the language allows finer distinctions between awareness structures and impossible-worlds structures.

In probabilistic awareness structures, the axioms of probability described by FHM are all valid. For example, $\ell(\varphi) = \ell(\psi)$ is valid in probabilistic awareness structures if $\varphi$ and $\psi$ are equivalent formulas. Using arguments similar to those in

Theorem 3.5, we can show that $\neg K \neg \ell(\varphi) = \ell(\psi)$ is valid in probabilistic awareness structures. Similarly, since $\ell(\varphi) + \ell(\neg\varphi) = 1$ is valid in probability structures, $\neg K(\neg(\ell(\varphi) + \ell(\neg\varphi) = 1))$ is valid in probabilistic awareness structures.

We can characterize properties of knowledge and likelihood in probabilistic awareness structures axiomatically. Let *Prob* denote a substitution instance of a valid formula in probabilistic logic (using the FHM axiomatization). By the observation above, *Prob* is sound in probabilistic awareness structures. Our reasoning has to take this into account. There is also an axiom *KL* that connects knowledge and likelihood:

$$K\varphi \Rightarrow \ell(\varphi) > 0. \tag{$KL$}$$

Let $\mathrm{AX}^P_{Ver}$ denote the axiom system consisting of $\{Prop, MP, Prob, KL, Ver\}$. Let $DC^P$ be the following strengthening of *DC*, somewhat in the spirit of *KC*:

$$\begin{aligned}&(K\varphi_1 \wedge \cdots \wedge K\varphi_n) \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)\\ &\quad \text{if } \mathrm{AX}^P_{Ver} \vdash \varphi_1 \wedge \cdots \wedge \varphi_n \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)\\ &\quad \text{and } \psi_1, \dots, \psi_m \text{ are likelihood formulas.}\end{aligned} \tag{$DC^P$}$$

Finally, even though *Ver* is not sound in KD45 probabilistic awareness structures, a weaker version, restricted to likelihood formulas, is sound, since there is a single probability distribution in probabilistic awareness structures. Let *WVer* be the following axiom:

$$K\varphi \Rightarrow \varphi \quad \text{if } \varphi \text{ is a likelihood formula.} \tag{$WVer$}$$

Let $\mathrm{AX}^P_{DC} = \{Prop, MP, Prob, DC^P, WVer, KL\}$ be the axiom system obtained by replacing *DC* in $\mathrm{AX}_{DC}$ by $DC^P$ and adding *Prob*, *WVer*, and *KL*.

**Theorem 4.1.**

(a) *$AX^P_{DC}$ is a sound and complete axiomatization of $\mathcal{L}^{K,QU}$ with respect to KD45 probabilistic awareness structures.*
(b) *$AX^P_{Ver}$ is a sound and complete axiomatization of $\mathcal{L}^{K,QU}$ with respect to S5 probabilistic awareness structures.*

**Proof.** (a) We first prove soundness. We have already argued that *Prob* is sound in KD45 probabilistic awareness structures. It is easy to see that *KL* is sound: let $M = (W, W', \pi, \mathcal{A}, \mu)$ be a KD45 probabilistic awareness structure, and let $w$ be a world in $W$ such that $(M, w) \models K\varphi$. This means that $\varphi$ is true at every world $w' \in W'$, and therefore, $\mu([\![\varphi]\!]_M \cap W') = \mu(W') > 0$, that is, $(M, w) \models \ell(\varphi) > 0$. Similarly, *WVer* is sound: let $M = (W, W', \pi, \mathcal{A}, \mu)$ be a KD45 probabilistic awareness structure, and let $w$ be a world in $W$ such that $(M, w) \models K\varphi$, with $\varphi$ a likelihood formula. This means that $\varphi$ is true at every world $w' \in W'$, and because $\varphi$ is a likelihood formula, the truth of $\varphi$ does not depend on the world. Thus, if $\varphi$ is true at some world, it is true at every world; in particular, it is true at $w$, so that $(M, w) \models \varphi$, as required. Finally, we show soundness of $DC^P$, using an argument similar to that in the proof of Theorem 3.6. Suppose that $M = (W, W', \pi, \mathcal{A}, \mu)$ is a KD45 probabilistic awareness structure, $w \in W$, $\mathrm{AX}^P_{Ver} \vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)$, for likelihood formulas $\psi_1, \dots, \psi_m$, and $(M, w) \models K\varphi_1 \wedge \cdots \wedge K\varphi_n$. Thus, $(M, w'') \models \varphi_1 \wedge \cdots \wedge \varphi_n$ for all $w'' \in W'$. But since each world in $W'$ is a model of $\mathrm{AX}^P_{Ver}$, if $w' \in W'$, we must have $(M, w') \models \psi_1 \vee \cdots \vee \psi_m$. Since $W' \neq \varnothing$, let $w'$ be an element of $W'$. For some $j \in \{1, \dots, m\}$, we must have $(M, w') \models \psi_j$. Because $\psi_j$ is a likelihood formula, and therefore its truth does not depend on the world, if $\psi_j$ is true at some world, then $\psi_j$ is true at every world. In particular, $(M, w) \models \psi_j$, and it follows that $(M, w) \models \psi_1 \vee \cdots \vee \psi_m$, as desired.

Completeness follows from combining techniques from the FHM completeness proof with those of Theorem 3.6. We briefly sketch the main ideas here. Define $Sub_P(\varphi)$ to be the least set containing $\varphi$, closed under subformulas, and containing $\ell(\psi) > 0$ if it contains a propositional formula $\psi$. It is easy to see that $|Sub_P(\varphi)| \leqslant 2|\varphi|$. Suppose that $\varphi$ is consistent with $\mathrm{AX}^P_{DC}$. Let $F$ be a maximal $\mathrm{AX}^P_{DC}$-consistent subset of $Sub_P(\varphi)$ that includes $\varphi$. Let $S$ consist of all truth assignments to primitive propositions. Using techniques of FHM, we can show that there must be a probability measure $\mu$ on $S$ that makes all the likelihood formulas in $F$ true. We remark for future reference that the FHM proof shows that we can take the set of truth assignments which get positive probability to be polynomial in the size of $|\varphi|$, and we can assume that the probability is rational, with a denominator whose size is polynomial in $|\varphi|$.

Let $H = \{\psi \mid K\psi \in F\} \cup \{\psi \mid \psi \in F, \psi \text{ is a likelihood formula}\}$. Arguments almost identical to those in Theorem 3.6 show that $H$ must be $\mathrm{AX}^P_{DC}$-consistent. Hence there is a maximal $\mathrm{AX}^P_{DC}$-consistent subset $F'$ of $Sub_P(\varphi)$ that contains $H$. We now construct a KD45 awareness structure $(\{w\} \cup W', W', \mathcal{A}, \mu')$ as follows. There is a world $w_v$ in $W'$ corresponding to each truth assignment $v$ such that $\mu(v) > 0$ and a world $w'$ corresponding to $F'$; we define $\mu'$ on $W'$ so that $\mu'(w') = 0$ and $\mu'(w_v) = \mu(v)$. Define $\pi$ so that $\pi(w_v) = v$, $\pi(w)(p) = \textbf{true}$ iff $p \in F$ and $\pi(w')(p) = \textbf{true}$ iff $p \in F'$. Finally, define $\mathcal{A}$ so that $\mathcal{A}(w_v) = \varnothing$, $\mathcal{A}(w') = \{\psi \mid K\psi \in F'\}$ and $\mathcal{A}(w) = \{\psi \mid K\psi \in F\}$. Now the same ideas as in the proof of Theorem 3.6 show that, for each formula $\psi \in Sub_P(\varphi)$ we have that $(M, w') \models \psi$ iff $\psi \in F'$ and $(M, w) \models \psi$ iff $\psi \in F$. Thus, $(M, w) \models \varphi$.

(b) The soundness of *Ver* in S5 probabilistic awareness structures follows easily by induction on the structure of $\varphi$ in $K\varphi$, using the fact that *WVer*—the special case of *Ver* when $\varphi$ is a likelihood formula—is sound in probabilistic awareness structures, and the argument for the soundness of *Ver* in S5 awareness structures.

The proof of completeness is similar in spirit to the proof of completeness in part (a); the modifications required are exactly those needed to prove Theorem 3.6(c). We leave details to the reader. □

Things change significantly when we move to probabilistic impossible-worlds structures. In particular, *Prob* is no longer sound. For example, even if $\varphi \Leftrightarrow \psi$ is valid, $\ell(\varphi) = \ell(\psi)$ is not valid, because we can have an impossible possible world with positive probability where both $\varphi$ and $\neg\psi$ are true. Similarly, $\ell(\varphi) + \ell(\neg\varphi) = 1$ is not valid. Indeed, both $\ell(\varphi) + \ell(\neg\varphi) > 1$ and $\ell(\varphi) + \ell(\neg\varphi) < 1$ are satisfiable in impossible-worlds structures: the former requires that there be an impossible possible world that gets positive probability where both $\varphi$ and $\neg\varphi$ are true, while the latter requires an impossible possible world with positive probability where neither is true. As a consequence, it is not hard to show that both $K\neg(\ell(\varphi) = \ell(\psi))$ and $K(\neg(\ell(\varphi) + \ell(\neg\varphi) = 1))$ are satisfiable in such impossible-worlds structures.[5] In fact, the only constraint on probability in probabilistic impossible-worlds structures is that it must be between 0 and 1. This constraint is expressed by the following axiom *Bound*:

$$\ell(\varphi) \geqslant 0 \wedge \ell(\varphi) \leqslant 1. \qquad (Bound)$$

We can characterize properties of knowledge and likelihood in probabilistic impossible-worlds structures axiomatically. Let $\mathrm{AX}^B_{imp} = \{Prop, MP, Ineq, Bound, KL, WVer\}$. We can think of $\mathrm{AX}^B_{imp}$ as being the core of probabilistic reasoning in impossible-worlds structures.

Let $\mathrm{AX}^B_{Ver}$ denote the axiom system consisting of $\{Prop, MP, Ineq, Bound, Ver, KL\}$. Let $KC^P$ denote the following extension of *KC*:

$$\begin{aligned}&(K\varphi_1 \wedge \cdots \wedge K\varphi_n) \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)\\ &\quad \text{if } \mathrm{AX}^P_{Ver} \vdash \varphi_1 \wedge \cdots \wedge \varphi_n \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)\\ &\quad \text{and } \psi_j \text{ is either a likelihood formula or of the form } K\psi', \text{ for } j = 1, \ldots, m. \qquad (KC^P)\end{aligned}$$

Here again, $DC^P$ is a special case of $KC^P$. Let $\mathrm{AX}^B_{KC} = \{Prop, MP, Ineq, Bound, KC^P, WVer, KL\}$ obtained by replacing *KC* in $\mathrm{AX}_{KC}$ by $KC^P$ and adding *Ineq*, *Bound*, *WVer* and *KL*.

**Theorem 4.2.**

(a) *$AX^B_{imp}$ is a sound and complete axiomatization of $\mathcal{L}^{K,QU}$ with respect to KD45$^-$ probabilistic impossible-worlds structures.*
(b) *$AX^B_{KC}$ is a sound and complete axiomatization of $\mathcal{L}^{K,QU}$ with respect to KD45 probabilistic impossible-worlds structures.*
(c) *$AX^B_{Ver}$ is a sound and complete axiomatization of $\mathcal{L}^{K,QU}$ with respect to S5 probabilistic impossible-worlds structures.*

**Proof.** (a) We first prove soundness. The argument is similar to the argument for soundness in Theorem 4.1. That *KL* and *WVer* are sound in probabilistic impossible-worlds structures follows from the same argument as in Theorem 4.1. To show that *Bound* is sound, note that for any probabilistic impossible-worlds structure $M$, $[\![\varphi]\!]_M \cap W' \subseteq W'$, so that $0 \leqslant \mu([\![\varphi]\!]_M) \leqslant 1$. Because this is independent of the actual world, $(M, w) \models \ell(\varphi) \geqslant 0 \wedge \ell(\varphi) \leqslant 1$ holds.

For completeness, given a formula $\varphi$ consistent with $\mathrm{AX}_{imp}$, let $F$ be a maximal $\mathrm{AX}_{imp}$-consistent subset of $Sub_P(\varphi)$ that includes $\varphi$. Consider the basic likelihood formulas in $F$. From these, we can get a system of linear inequalities by replacing each term $\ell(\psi)$ by a variable $x_\psi$. We add an inequality $0 \leqslant x_\psi \leqslant 1$ for each formula $\psi \in Sub_P(\varphi)$. Using the arguments of FHM, we can show that this set of inequalities must be satisfiable (otherwise $F$ would not be $\mathrm{AX}_{imp}$-consistent). Take a solution. Without loss of generality, we have subformulas listed so that $x_{\psi_1} \leqslant x_{\psi_2} \leqslant \cdots \leqslant x_{\psi_n}$. Let $n^*$ be the least $m$ such that $x_{\psi_m} = 1$; if $x_{\psi_n} < 1$, then let $n^* = n + 1$. Consider a probabilistic impossible-worlds structure $(\{w\}, \{w_1, \ldots, w_{n+1}, w\}, \pi, \mathcal{C}, \mu)$, where we define $\pi$, $\mathcal{C}$ and $\mu$ as follows:

- $\pi(w)(p) = \mathbf{true}$ iff $p \in F$;
- $\mu(w_1) = x_{\psi_1}$, $\mu(w_j) = x_{\psi_j} - x_{\psi_{j-1}}$ for $j = 2, \ldots, n$, and $\mu(w_{n+1}) = 1 - \mu(w_n)$;
- $\mathcal{C}(w_j) = \{\psi_j, \ldots, \psi_n\}$ for $j = 1, \ldots, n^*$;
- $\mathcal{C}(w_j) = \mathcal{C}(w_{n^*})$ if $j = n^* + 1, \ldots, n + 1$.

We leave it to the reader to show that $(M, w) \models \varphi$.

(b) We show soundness of $KC^P$ with respect to KD45 probabilistic impossible-worlds structures. For suppose that $M = (W, W', \pi, \mathcal{C}, \pi)$ is a KD45 probabilistic impossible-worlds structure, $w \in W$, $\mathrm{AX}_{Ver} \vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \Rightarrow (\psi_1 \vee \cdots \vee \psi_m)$, where each $\psi_j$ either a likelihood formula or of the form $K\psi'$, and $(M, w) \models K\varphi_1 \wedge \cdots \wedge K\varphi_n$. Thus, $(M, w'') \models \varphi_1 \wedge \cdots \wedge \varphi_n$ for all $w'' \in W'$. But since each world in $W \cap W'$ is a model of $\mathrm{AX}^P_{Ver}$, if $w' \in W \cap W'$, we must have $(M, w') \models \psi_1 \vee \cdots \vee \psi_m$.

[5] We remark that Cozic [2], who considers the logical omniscience problem in the context of probabilistic reasoning, makes somewhat similar points. Although he does not formalize things quite the way we do, he observes that, in his setting, impossible-worlds structures seem more expressive than awareness structures.

Moreover, since $W \cap W' \neq \varnothing$, there must be some world $w' \in W \cap W'$. It follows that, for some $j \in \{1, \ldots, m\}$, $(M, w') \models \psi_j$. There are two cases. If $\psi_j$ is a likelihood formula, then its truth does not depend on the world, so that if $\psi_j$ is true at some world, then $\psi_j$ is true at every world. In particular, $(M, w) \models \psi_j$, and it follows that $(M, w) \models \psi_1 \vee \cdots \vee \psi_m$, as desired. If $\psi_j$ is a formula of the form $K\psi'$, then $(M, w'') \models \psi'$ for all $w'' \in W'$, so $(M, w) \models K\psi'$, that is, $(M, w) \models \psi_j$. It follows that $(M, w) \models \psi_1 \vee \cdots \vee \psi_m$, as desired.

The completeness argument is similar in spirit to that of part (a) and left to the reader.

(c) For soundness, as in the proof of Theorem 4.1, the soundness of *Ver* in S5 probabilistic impossible-worlds structures follows by induction on the structure of $\varphi$ in $K\varphi$.

The completeness argument is similar in spirit to that of part (a) and left to the reader. □

Observe that Theorem 4.2 is true even though probabilities are standard in impossible worlds: the probabilities of worlds still sum to 1. It is just the truth assignment to formulas that behaves in a nonstandard way in impossible worlds. Intuitively, while the awareness approach is modeling certain consequences of resource-boundedness in the context of knowledge, it does not do so for probability. On the other hand, the impossible-worlds approach seems to extend more naturally to accommodate the consequences of resource-boundedness in probabilistic reasoning; see Section 5 for more discussion of this issue.

**Corollary 4.3.** *The satisfiability problem for the language $\mathcal{L}^{K,QU}$ with respect to KD45 probabilistic awareness structures (resp., S5 probabilistic awareness structures, KD45$^-$ probabilistic impossible-worlds structures, KD45 probabilistic impossible-worlds structures, S5 probabilistic impossible-worlds structures) is NP-complete.*

**Proof.** Again, NP-hardness follows immediately from the observation that $\mathcal{L}^{K,QU}$ contains propositional logic. The fact that the satisfiability problem with respect to each of these classes of structures is in NP follows from the constructions of Theorems 4.1 and 4.2, which show that if a formula $\varphi$ is satisfiable with respect to KD45 probabilistic awareness structures (resp., S5 probabilistic awareness structures, KD45$^-$ probabilistic impossible-worlds structures, KD45 probabilistic impossible-worlds structures, S5 probabilistic impossible-worlds structures), then it is consistent with respect to $\mathrm{AX}^P_{DC}$ (resp., $\mathrm{AX}^P_{Ver}$, $AX^B_{imp}$, $AX^B_{KC}$, $AX^B_{Ver}$) which in turn implies that it is satisfiable in a KD45 probabilistic awareness structure (resp., S5 probabilistic awareness structure, KD45$^-$ probabilistic impossible-worlds structure, KD45 probabilistic impossible-worlds structure, S5 probabilistic impossible-worlds structure) $M = (W, W', \ldots)$ with a small number of worlds—polynomial in the length of $\varphi$ in each case. Just as in the proof of Corollary 3.7, without loss of generality, we can assume that, in the case of probabilistic awareness structures, at each world $w \in W$, $\mathcal{A}(w)$ is a subset of $Sub(\varphi)$, the set of subformulas of $\varphi$, and $\pi(w)(p) = \textbf{true}$ only if $p$ is a subformula of $\varphi$; similarly, in the case of probabilistic impossible-worlds structures, we can assume that for each impossible world $w'$, $\mathcal{C}(w')$ is a subset of the subformulas of $\varphi$. Finally, using the arguments of FHM, we can argue without loss of generality that the probability distributions $\mu$ are described in size polynomial in the length of $\varphi$. (The probability distributions in all structures can be taken to assign small—polynomial-size—rational probabilities to every world, where the size of a rational number is the sum of the sizes of the numerator and denominator when they are relatively prime.) Thus, we can guess a satisfying structure for $\varphi$ and verify that it satisfies $\varphi$ in time polynomial in the length of $\varphi$. □

## 5. Pragmatic issues

Even in settings where the four approaches are equi-expressive, they model lack of logical omniscience quite differently. We thus have to deal with different issues when attempting to use one of them in practice. By "in practice", we mean attempting to use one of the models above to capture a particular scenario about which one wants to reason—as opposed, say, to capturing a scenario using axioms in the logic, and reasoning exclusively using the proof rules of the logic. Issues like the following can arise: If we are using a syntactic structure to represent a given situation, we need to explain where the function $\mathcal{C}$ is coming from; with an awareness structure, we must explain where the awareness function is coming from; with an algorithmic knowledge structure, we must explain where the algorithm is coming from; and with an impossible-worlds structure, we must explain what the impossible worlds are.

There seem to be three quite distinct intuitions underlying the lack of logical omniscience. As we now discuss, these intuitions can guide the choice of approach, and match closely the solutions described above. We discuss, for each intuition, the extent to which each of the approaches to dealing with logical omniscience can capture that intuition. While the discussion in this section is somewhat informal, we believe that these observations will prove important when actually trying to decide how to model lack of logical omniscience in practice.

### 5.1. Lack of awareness

The first intuition is lack of awareness of some primitive notions: for example, an agent reasoning in 2002 about possible outcomes of an attack on Iraq may not have even contemplated outcomes like suicide bombers. If "suicide bombers cause many deaths" is a primitive proposition in the language of a more sophisticated modeler, then the agent would not be aware of this proposition, and hence could not even consider it possible that suicide bombers would cause many deaths. (Such reasoning becomes more interesting if there is more than one agent, and they are aware of different primitive propositions.)

This can be modeled reasonably well using an awareness structure where the awareness function is *generated by primitive propositions*. We assume that the agent is unaware of certain primitive propositions, and is unaware of exactly those formulas that contain a primitive proposition of which the agent is unaware. This intuition is quite prevalent in the economics community, and all the standard approaches to modeling lack of logical omniscience in the economics literature [22,23,3,16] can essentially be understood in terms of awareness structures where awareness is generated by primitive propositions [11,14].

Of course, how we choose these primitive propositions is critical to the modeling process. What counts as a primitive proposition is ultimately in the eye of the beholder, and different modelers may well choose different primitive propositions to capture a particular scenario. However, for many scenarios, the choice of primitive propositions is usually clear and uncontroversial, as witnessed by the popularity of the approach in the economics literature.

If awareness is generated by primitive propositions, constructing an awareness structure corresponding to a particular situation is no more (or less!) complicated that constructing a Kripke structure to capture knowledge without awareness. Determining the awareness sets for notions of awareness that are not generated by primitive propositions may be more complicated. It is also worth stressing that an awareness structure must be understood as the modeler's view of the situation. For example, if awareness is generated by primitive propositions and agent 1 is not aware of a primitive proposition $p$, then agent 1 cannot contemplate a world where $p$ is true (or false); in the model from agent 1's point of view, $p$ does not exist.

How do the other approaches fare in modeling lack of awareness? To construct a syntactic structure, we need to know all sentences that an agent knows before constructing the model. This may or may not be reasonable. But it does not help in discovering properties of knowledge in a given situation. As observed in [10], the syntactic approach is really only a representation of knowledge. Algorithmic knowledge can deal with lack of awareness reasonably well, provided that there is an algorithm $\mathbb{A}_a$ for determining what the agent is aware of and an algorithm $\mathbb{A}_k$ for determining whether a formula is true in every world in $W'$, the set of worlds that the agent considers possible. If so, given a query $\varphi$, the algorithmic approach would simply invoke $\mathbb{A}_a$ to check whether the agent is aware of $\varphi$; if so, then the agent invokes $\mathbb{A}_k$. For example, if awareness is generated by primitive propositions, then $\mathbb{A}_a$ is the algorithm that, given query $\varphi$, checks whether all the primitive propositions in $\varphi$ are ones the agent is aware of; and we can take $\mathbb{A}_k$ to be the algorithm that does model checking to see if $\varphi$ is true in every world of $W'$. (This can be done in time polynomial in $W'$; see [10].) In impossible-worlds structures, we can interpret lack of awareness of $\varphi$ as meaning that neither $\varphi$ nor $\neg\varphi$ is true at all worlds the agent considers possible. Thus, if there is any nontrivial lack of awareness, then all the worlds that the agent considers possible will be impossible worlds. However, these impossible worlds have a great deal of structure: we can require that for all the formulas $\varphi$ that the agent is aware of, exactly one of $\varphi$ and $\neg\varphi$ is true at each world the agent considers possible. As we observed earlier, an awareness structure must be viewed as the *modeler's* view of the situation. Arguably, the impossible-worlds structure better captures the agent's view.

### *5.2. Lack of computational ability*

The second intuition is computational: an agent simply might not have the resources to compute the required answer. But then the question is how to model this lack of computational ability. There are two cases of interest, depending on whether we have an explicit algorithm in mind. If we have an explicit algorithm, then it is relatively straightforward. For example, Konolige [18] uses a syntactic approach and gives an explicit characterization of $\mathcal{C}$ by taking it to be the set of formulas that can be derived from a fixed initial set of formulas by using a sound but possibly incomplete set of inference rules. Note that Konolige's approach makes syntactic knowledge an instance of algorithmic knowledge. (See also Pucella [27] for more details on knowledge algorithms given by inference rules.)

Algorithmic knowledge can be viewed as a generalization of Konolige's approach in this setting, since it allows for the possibility that the algorithm used by the agent to compute what he knows may not be easily expressible as a set of inference rules over formulas. For example, Berman, Garay, and Perry [1] implicitly use a particular form of algorithmic knowledge in their analysis of *Byzantine agreement* (this is the problem of getting all nonfaulty processes in a system to coordinate, despite the presence of failures). Roughly speaking, they allow agents to perform limited tests based on the information they have; agents know only what follows from these limited tests. But these tests are not characterized axiomatically. As shown by Halpern and Pucella [13], algorithmic knowledge is also a natural way to capture adversaries in security protocols.

**Example 5.1.** Security protocols are generally analyzed in the presence of an adversary that has certain capabilities for decoding the messages he intercepts. There are of course restrictions on the capabilities of a reasonable adversary. For instance, the adversary may not explicitly know that he has a given message if that message is encrypted using a key that the adversary does not know. To capture these restrictions, Dolev and Yao [4] gave a now-standard description of the capabilities of adversaries. Roughly speaking, a Dolev–Yao adversary can decompose messages, or decipher them if he knows the right keys, but cannot otherwise "crack" encrypted messages. The adversary can also construct new messages by concatenating known messages, or encrypting them with a known encryption key.

Algorithmic knowledge is a natural way to capture the knowledge of a Dolev–Yao adversary [13]. We can use a knowledge algorithm $\mathtt{A}^{\mathrm{DY}}$ to compute whether the adversary can *extract* a message $m$ from a set $H$ of messages that he has intercepted, where the extraction relation $H \vdash_{DY} m$ is defined by following inference rules:

$$\frac{m \in H}{H \vdash_{DY} m} \quad \frac{H \vdash_{DY} \{m\}_k \quad H \vdash_{DY} k}{H \vdash_{DY} m} \quad \frac{H \vdash_{DY} m_1 \cdot m_2}{H \vdash_{DY} m_1} \quad \frac{H \vdash_{DY} m_1 \cdot m_2}{H \vdash_{DY} m_2},$$

where $m_1 \cdot m_2$ is the concatenation of messages $m_1$ and $m_2$, and $\{m\}_k$ is the encryption of message $m$ with key $k$.

The knowledge algorithm $\mathtt{A}^{\mathrm{DY}}$ simply implements a search for the derivation of a message $m$ from the messages that the adversary has received and the initial set of keys, using the inference rules above. More precisely, we assume the language has formulas $has(m)$, interpreted as "the agent possesses message $m$". When queried for a formula $has(m)$, the knowledge algorithm $\mathtt{A}^{\mathrm{DY}}$ simply checks if $H \vdash_{DY} m$, where $H$ is the set of messages intercepted by the adversary. Thus, the formula $K(has(m))$, which is true if and only if $\mathtt{A}^{\mathrm{DY}}$ says "Yes" to query $has(m)$, that is, if and only if $H \vdash_{DY} m$, says that the adversary can extract $m$ from the messages he has intercepted.

However, even when our intuition is computational, at times the details of the algorithm do not matter (and, indeed, may not be known to the modeler). In this case, awareness may be more useful than algorithmic knowledge.

**Example 5.2.** Suppose that Alice is trying to reason about whether or not an eavesdropper Eve has managed to decrypt a certain message. The intuition behind Eve's inability to decrypt is computational, but Alice does not know which algorithm Eve is using. An algorithmic knowledge structure is typically appropriate if there are only a few algorithms that Eve might be using, and her ability to decrypt depends on the algorithm.[6] On the other hand, Alice might have no idea of what Eve's algorithm is, and might not care. All that matters to her analysis is whether Eve has managed to decrypt. In this case, using a syntactic structure or an awareness structure seems more appropriate. Suppose that Alice wants to model her uncertainty regarding whether Eve has decrypted the message. She could then use an awareness structure with some possible worlds where Eve is aware of the message, and others where she is not, with the appropriate probability on each set. Alice can then reason about the likelihood that Eve has decrypted the message without worrying about how she decrypted it.

What about the impossible-worlds approach? It cannot directly represent an algorithm, of course. However, if there is algorithm $\mathtt{A}$ that characterizes an agent's computational process, then we can simply take $W' = \{w'\}$ and define $\mathcal{C}(w') = \{\varphi \mid \mathtt{A}(\varphi) = \text{"Yes"}\}$. Indeed, we can give a general computational interpretation of the impossible-worlds approach. The worlds $w$ such that $\mathcal{C}(w)$ are precisely those worlds where the algorithm answers "Yes" when asked about $\varphi$. If neither $\varphi$ nor $\neg\varphi$ is in $\mathcal{C}(w)$, that just means that the algorithm was not able to determine whether $\varphi$ was true or false. If the algorithm answers "Yes" to both $\varphi$ and $\neg\varphi$, then clearly the algorithm is not sound, but it may nevertheless describe how a resource-bounded agent works.

This intuition also suggests how we can model the lack of computational ability illustrated by Example 5.2 using impossible worlds. Suppose that we add new primitive propositions of the form $cont(m) = c$ to the language that say that the content of a message $m$ is $c$. Then in a world where Alice cannot decrypt $c$, neither $cont(m) = c$ nor $\neg(cont(m) = c)$ would be true.

### *5.3. Imperfect understanding of the model*

Sometimes an agent's lack of logical omniscience is best thought of as stemming from "mistakes" in constructing the model (which perhaps are due to lack of computational ability).

**Example 5.3.** Suppose that Alice does not know whether a number $n$ is prime. Although her ignorance regarding $n$'s primality can be viewed as computationally based—given enough time and energy, she could in principle figure out whether $n$ is prime—she need not be using a specific algorithm to compute her knowledge (at least, not one that she can easily describe). Nor can her state of mind be modeled in a natural way using an awareness structure or a syntactic structure. Intuitively, there should at least two worlds she considers possible, one where $n$ is prime, and one where $n$ is not. However, $n$ is either prime or it is not. If $n$ is actually prime, then there cannot be a possible world where $n$ is not prime; similarly, if $n$ is composite, there cannot be a possible world where $n$ is prime. This problem can be modeled naturally using impossible worlds. Now there is no problem having a world where $n$ is prime (which is an impossible world if $n$ is actually composite) and a world where $n$ is composite (which is an impossible world if $n$ is actually prime). In this structure, it is also seems reasonable to assume that Alice knows that she does not know that $n$ is prime (so that the formula $\neg K Prime_n$ is true even in the impossible worlds).

It is instructive to compare this with the awareness approach. Suppose that $n$ is indeed prime and an external modeler knows this. Then he can describe Alice's state of mind with one world, where $n$ is prime, but Alice is not aware that $n$ is

[6] What is required here is an algorithmic knowledge structure with two agents. There will then be different algorithms for Eve associated with different states. We omit here the straightforward details of how this can be done; see [12].

prime. Thus, $\neg K Prime_n$ holds at this one world. But note that this is not because Alice considers it possible that $n$ is not prime; rather, it is because Alice cannot determine whether $n$ is prime using her internal algorithm. If Alice is aware of the formula $\neg K Prime_n$ at this one world, then $K \neg K Prime_n$ also holds. Again, we should interpret this as saying that Alice knows that she cannot determine whether $n$ is prime using her internal algorithm.

The impossible-worlds approach seems like a natural one in Example 5.3 and many other settings. As we saw, awareness in this situation does not quite capture what is going on here. Algorithmic knowledge fares somewhat better, but it would require us to have a specific algorithm in mind; in Example 5.3, this would force us to interpret "knows that a number is prime" as "knows that a number is prime as tested by a particular factorization algorithm".

The impossible-worlds approach can sometimes be difficult to apply, however, because it is not always clear what impossible worlds to incorporate in a model. While there has been a great deal of discussion (particularly in the philosophy literature) concerning the "metaphysical status" of impossible worlds (cf. [30]), the pragmatics of generating impossible worlds has received comparatively little attention. Hintikka [17] argues that Rantala's [28] urn models are suitable candidates for impossible worlds. In decision theory, Lipman [19] uses impossible-worlds structures to represent the preferences of an agent who may not be able to distinguish logically equivalent outcomes; the impossible worlds are determined by the preference order. None of these approaches address the problem of generating the impossible worlds even in a simple example such as Example 5.3, especially if the worlds have some structure.

We view impossible worlds as describing the agent's subjective view of a situation. The modeler may know that these impossible worlds are truly impossible, but the agent does not. In many cases, the intuitive reason that the agent does not realize that the impossible worlds are in fact impossible is that the agent does not look carefully at the worlds. Consider Example 5.3. Let $Prime_n$, for various choices of $n$, be a primitive proposition saying that the number $n$ is prime. Suppose that the worlds are models of arithmetic, which include as domain elements the natural numbers with multiplication defined on them. If $Prime_n$ is interpreted as being true in a world when there do not exist numbers $n_1$ and $n_2$ in that world such that $n_1 \times n_2 = n$, then how does the agent conceive of the impossible worlds? If the agent were to look carefully at a world where $Prime_n$ holds, he might realize that there are in fact two numbers $n_1$ and $n_2$ such that $n_1 \times n_2 = n$. But if $n$ is not prime, how do we capture the fact that the agent "mistakenly" constructed a world where there are numbers $n_1$ and $n_2$ such that $n_1 \times n_2 = n$ if we also assume that the agent understands basic multiplication?

We now sketch a new approach to constructing an impossible-worlds structure that seems appropriate for such examples. The approach is motivated by the observation that the set of worlds in a Kripke structure is explicitly specified, as is the truth assignment on these worlds. Introspectively, this is not the way in which we model situations. Instead, the set of possible worlds is described implicitly, as is the interpretation $\pi$, as the set of worlds satisfying some condition.[7] This set of worlds may well include some impossible worlds. The impossible-worlds structure corresponding to a situation, therefore, is made up of all worlds satisfying the implicit description, perhaps refined so that "clearly impossible" worlds are not considered. What makes a world clearly impossible should be determined by a simple test; for example, such a simple test might determine that 3 is prime, but would not be able to determine that $2^{24036583} - 1$ is prime.

We can formalize this construction as follows. An implicit structure is a tuple $I = (S, T, \mathcal{C})$, where $S$ is a set of possible worlds, $T$ is a filter on worlds (a test on worlds that returns either **true** or **false**), and $\mathcal{C}$ associates with every world in $S$ a set (possibly inconsistent) of propositional formulas. Test $T$ returns **true** for every world in $S$ that the agent considers possible. An implicit structure $I = (S, T, \mathcal{C})$ induces an impossible-worlds structure $M_I = (W, W', \pi, \mathcal{C})$ given by:

$$\begin{aligned}
&W = \{ w \in S \mid \mathcal{C}(w) \text{ is consistent} \}, \\
&W' = \{ w \in S \mid T(w) = \textbf{true} \}, \\
&\pi(w) = \mathcal{C}(w)|_{\Phi} \quad \text{for } w \in W, \\
&\mathcal{C} = \mathcal{C}|_{(W'-W)}.
\end{aligned}$$

We can refine the induced impossible-worlds structure by allotting more resources to test $T$. Intuitively, as an agent performs more introspection, she can recognize more worlds as being impossible. (Manne [20] investigates a related approach, using a temporal structure at each world to capture the evolution of knowledge as the agent introspects over time.)

Consider the primality example again. The agent is likely to care about the primality of only a few numbers, say $n_1, \ldots, n_k$. Let $\Phi = \{Prime_{n_1}, \ldots, Prime_{n_k}\}$. The agent's inability to compute whether $n_1, \ldots, n_k$ are prime is described implicitly by having worlds where any combination of them is prime. The details of how multiplication works in a world is not specified in the implicit description. Thus, the implicit structure $I = (S, T, \mathcal{C})$ corresponding to this description will have $S$ consisting of $2^k$ worlds, where each world is a standard model of arithmetic together with a truth assignment to the primitive propositions in $\Phi$. The set of formulas $\mathcal{C}(w)$ consists of all propositional formulas true under the truth assignment at $w$. The agent realizes that all but one of these worlds is impossible, but cannot compute which one is the possible world. Thus, we take $T(w) = \textbf{true}$ for all worlds $w$. Of course, after doing some computation, the agent may realize that, say, $n_1$ is

[7] In multiagent settings, where the worlds that the agent considers possible are defined by an accessibility relation, we expect the accessibility relation to be described implicitly as well.

prime and $n_2$ is composite. The agent would then refine the model by taking $T$ to consider possible only worlds in which $n_1$ is prime and $n_2$ is composite.

The use of an implicit description as a recipe for constructing possible (and impossible) worlds is quite general, as the following example illustrates.

**Example 5.4.** Suppose that we have a database of implications: rules of the form $C_1 \Rightarrow C_2$, where $C_1$ and $C_2$ are conjunctions of literals—primitive propositions and their negations. Suppose that the vocabulary of the conclusions of these rules is disjoint from the vocabulary of the antecedents. This is a slight simplification of, for example, digital rights management policies, where the conclusion typically has the form $\mathit{Permitted}(a, b)$ or $\neg\mathit{Permitted}(a, b)$ for some agent $a$ and action $b$, and *Permitted* is not allowed to appear in the antecedent of rules [15]. Rather than explicitly constructing the worlds compatible with the rules, a user might construct a naive implicit description of them. More specifically, suppose that we have a finite set of agents, say $a_1, \ldots, a_n$, and a finite set of actions, say $b_1, \ldots, b_m$. Consider the implicit structure $I = (S, T, \mathcal{C})$, where each world $w$ in $S$ is a truth assignment to the atomic formulas that appear in the antecedents of rules, augmented with all the literals in the conclusions of rules whose antecedent is true in $w$; furthermore, take $T(w) = \textbf{true}$ for all $w \in S$, and $\mathcal{C}(w)$ to be all propositional formulas true under the truth assignment at world $w$. Thus, for example, if a rule says $\mathit{Student}(a) \wedge \mathit{Female}(a) \Rightarrow \mathit{Permitted}(a, \mathit{Play\text{-}sports})$, then in a world where $\mathit{Student}(a)$ and $\mathit{Female}(a)$ are true, then so is $\mathit{Permitted}(a, \mathit{Play\text{-}sports})$. Similarly, if we have a rule that says $\mathit{Faculty}(a) \wedge \mathit{Female}(a) \Rightarrow \neg\mathit{Permitted}(a, \mathit{Play\text{-}sports})$, then in a world where $\mathit{Faculty}(a)$ and $\mathit{Female}(a)$ are true, $\neg\mathit{Permitted}(a, \mathit{Play\text{-}sports})$ as well. Of course, in a world $\mathit{Faculty}(a)$, $\mathit{Student}(a)$, and $\mathit{Female}(a)$ are all true, both $\mathit{Permitted}(a, \mathit{Play\text{-}sports})$ and $\neg\mathit{Permitted}(a, \mathit{Play\text{-}sports})$ are true; this is an impossible world. This type of implicit description (and hence, impossible-worlds structure) should also be useful for characterizing large databases, when it is not possible to list all the tables explicitly.

## 6. Conclusion

Many solutions have been proposed to the logical omniscience problem, differing as to the intuitions underlying the lack of logical omniscience. There has been comparatively little work on comparing approaches. We have attempted to fill this gap here, focusing on two aspects, expressiveness and pragmatics, for four popular approaches.

In comparing the expressive power of the approaches, we started with the well-known observation that the approaches are equi-expressive in the propositional case. However, this observation is true only if we allow the agent not to consider any world possible. If we require that at least one world be possible, then we get a difference in expressive power. This is particularly relevant when we have probabilities, because there has to be at least one world over which to assign probability. Indeed, when considering logical omniscience in the presence of probability, there can be quite significant differences in expressive power between the approaches, particularly awareness and impossible worlds.

Considering the pragmatics of logical omniscience, we identified some guiding principles for choosing an approach to model a situation, based on the source of the lack of logical omniscience in that situation. As we show, coming up with an appropriate structure can be nontrivial. We illustrate a general approach to deriving an impossible-worlds structure based on an implicit description of the situation, which seems to be appropriate for a number of situations of interest. Our discussion suggests that the impossible-worlds approach may be particularly appropriate for representing an agent's subjective view of the world.

## References

[1] P. Berman, J. Garay, K.J. Perry, Towards optimal distributed consensus, in: Proc. 30th IEEE Symposium on the Foundations of Computer Science (FOCS'89), 1989, pp. 410–415.
[2] M. Cozic, Impossible states at work: Logical omniscience and rational choice, in: Cognitive Economics: New Trends, in: Contrib. Econ. Anal., vol. 280, Elsevier, 2007, Ch. 2, pp. 47–68.
[3] E. Dekel, B. Lipman, A. Rustichini, Standard state-space models preclude unawareness, Econometrica 66 (1998) 159–173.
[4] D. Dolev, A.C. Yao, On the security of public key protocols, IEEE Trans. Inform. Theory 29 (2) (1983) 198–208.
[5] R.A. Eberle, A logic of believing, knowing and inferring, Synthese 26 (1974) 356–382.
[6] H.B. Enderton, A Mathematical Introduction to Logic, Academic Press, 1972.
[7] R. Fagin, J.Y. Halpern, Belief, awareness, and limited reasoning, Artificial Intelligence 34 (1988) 39–76.
[8] R. Fagin, J.Y. Halpern, Reasoning about knowledge and probability, J. ACM 41 (2) (1994) 340–367.
[9] R. Fagin, J.Y. Halpern, N. Megiddo, A logic for reasoning about probabilities, Inform. Comput. 87 (1/2) (1990) 78–128.
[10] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Reasoning about Knowledge, MIT Press, 1995.
[11] J.Y. Halpern, Alternative semantics for unawareness, Games Econ. Behav. 37 (2001) 321–339.
[12] J.Y. Halpern, Y. Moses, M.Y. Vardi, Algorithmic knowledge, in: Proc. 5th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'94), Morgan Kaufmann, 1994, pp. 255–266.
[13] J.Y. Halpern, R. Pucella, Modeling adversaries in a logic for reasoning about security protocols, in: Proc. Workshop on Formal Aspects of Security (FASec'02), in: Lecture Notes in Comput. Sci., vol. 2629, 2002, pp. 115–132.
[14] J.Y. Halpern, L.C. Rêgo, Interactive unawareness revisited, in: Theoretical Aspects of Rationality and Knowledge: Proc. Tenth Conference (TARK 2005), 2005, pp. 78–91.
[15] J.Y. Halpern, V. Weissman, Using first-order logic to reason about policies, ACM Trans. Inform. Syst. Secur. 11 (4) (2008) 1–41.
[16] A. Heifetz, M. Meier, B. Schipper, Interactive unawareness, J. Econ. Theory 130 (2006) 78–94.
[17] J. Hintikka, Impossible possible worlds vindicated, J. Philos. Logic 4 (1975) 475–484.
[18] K. Konolige, A Deduction Model of Belief, Morgan Kaufmann, San Francisco, 1986.

[19] B.L. Lipman, Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality, Rev. Econ. Stud. 66 (2) (1999) 339–361.
[20] K. Manne, Towards a solution to the problem of logical omniscience for resource-bounded agents: An impossible worlds approach to temporal epistemic logic, unpublished manuscript; an abstract appears in: Proc. 2005 Conference of the Australasian Association of Philosophy, 2005.
[21] J. McCarthy, M. Sato, T. Hayashi, S. Igarishi, On the model theory of knowledge, Tech. Rep. STAN-CS-78-657, Stanford University, 1979.
[22] S. Modica, A. Rustichini, Awareness and partitional information structures, Theory Dec. 37 (1994) 107–124.
[23] S. Modica, A. Rustichini, Unawareness and partitional information structures, Games Econ. Behav. 27 (2) (1999) 265–298.
[24] R.C. Moore, G. Hendrix, Computational models of beliefs and the semantics of belief sentences, Technical Note 187, SRI International, Menlo Park, CA, 1979.
[25] A. Moreno, Avoiding logical omniscience and perfect reasoning: A survey, AI Commun. 11 (2) (1998) 101–122.
[26] Y. Moses, M.R. Tuttle, Programming simultaneous actions using common knowledge, Algorithmica 3 (1988) 121–169.
[27] R. Pucella, Deductive algorithmic knowledge, J. Logic Comput. 16 (2) (2006) 287–309.
[28] V. Rantala, Urn models: A new kind of non-standard model for first-order logic, J. Philos. Logic 4 (1975) 455–474.
[29] V. Rantala, Impossible worlds semantics and logical omniscience, Acta Philosophica Fennica 35 (1982) 18–24.
[30] R. Stalnaker, Impossibilities, Philos. Topics 24 (1996) 193–204.
[31] H. Wansing, A general possible worlds framework for reasoning about knowledge and belief, Studia Logica 49 (4) (1990) 523–539.