Dealing With Logical Omniscience

Joseph Y. Halpern
Cornell University
Ithaca, NY 14853 USA
halpern@cs.cornell.edu

Riccardo Pucella
Northeastern University
Boston, MA 02115 USA
riccardo@ccs.neu.edu

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 they are not. In particular, adding probabilities to the language allows for finer distinctions between different approaches.

1 Introduction

Logics of knowledge based on possible-world semantics are 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 φ if φ 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 literatureas 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) [Moses and Tuttle 1988]; 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 (e.g., 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 [Fagin, Halpern, Moses, and Vardi 1995, Chapter 9] and [Moreno 1998]). A far from exhaustive list of approaches includes:

- *syntactic approaches* [Eberle 1974; Moore and Hendrix 1979; Konolige 1986], where an agent's knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);

- *awareness* [Fagin and Halpern 1988], where an agent knows φ if she is aware of φ and φ is true in all the worlds she considers possible;
- *algorithmic knowledge* [Halpern, Moses, and Vardi 1994] where, roughly speaking, an agent knows φ if her knowledge algorithm returns "Yes" on a query of φ; and
- *impossible worlds* [Rantala 1982], where the agent may consider possible worlds that are logically inconsistent (for example, where p and $\neg p$ are both true).

Which approach is best to use, of course, depends on the application. Our goal is to elucidate the aspects of the application that make a logic more or less appropriate. We focus here on the expressive power of these approaches. 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 [Wansing 1990; Fagin, Halpern, Moses, and Vardi 1995] 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 is particularly relevant with probability in the picture.

Expressive power is only part of the story. In the full version of this paper [Halpern and Pucella 2007], we consider (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. Also for reasons of space, proofs of our technical results have been omitted, and can be found in the full paper.

2 The Four Approaches: A Review

We now review the standard possible-worlds approach and the four approaches to dealing 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

Starting with a set Φ of propositional formulas, we close off under conjunction, negation, and the K operator. Call the resulting language $\mathcal{L}^K$. We give semantics to these formulas using Kripke structures. For simplicity, we focus on approaches that satisfy the K45 axioms (as well as KD45 and S5). In this case, a *K45 Kripke structure* is a triple (W, W', π), 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 π is an *interpretation* that associates with each world a truth assignment $\pi(w)$ to the primitive propositions in Φ. 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 $\pi(w)(p) = \mathbf{true}$ if $p \in \Phi$.

$(M, w) \models \neg\varphi$ 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 $(K\varphi \wedge K(\varphi \Rightarrow \psi)) \Rightarrow K\psi$, which says that an agent's knowledge is closed under implication. In addition, the *knowledge generalization* inference rule is sound: From φ infer $K\varphi$. 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, the formula $K(\mathit{false})$ is true. A *KD45 Kripke structure* is a K45 Kripke structure (W, W', π) where $W' \neq \varnothing$. 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', π) 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

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

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 φ, not only does φ have to be true at all the worlds she considers possible, but she has to be aware of φ as well. A *K45 awareness structure* is a tuple $M = (W, W', \pi, \mathcal{A})$, where (W, W', π) is a K45 Kripke structure and $\mathcal{A}$ maps worlds to sets of formulas. We now define

$(M, w) \models K\varphi$ iff $(M, w') \models \varphi$ for all $w' \in W'$ and $\varphi \in \mathcal{A}(w)$.[1]

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', π) is a KD45 structure, and an S5 awareness structure when (W, W', π) 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', π) is a K45 Kripke structure and $\mathtt{A}$ is a *knowledge algorithm* that returns "Yes", "No", or "?" given a formula φ.[2] Intuitively, $\mathtt{A}(\varphi)$ returns "Yes" if the agent can compute that φ is true, "No" if the agent can compute that φ is false, and "?" otherwise. In algorithmic knowledge structures,

$(M, w) \models K\varphi$ iff $\mathtt{A}(\varphi) =$ "Yes".

An important class of knowledge algorithms consists of the *sound* knowledge algorithms. When a sound knowledge algorithm returns "Yes" to a query φ, then the agent knows (in the standard sense) φ, and when it returns "No" to a query φ, then the agent does not know (again, in the standard sense) φ. 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 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.)

Algorithmic knowledge can be seen as a generalization of a number of approaches in the literature, although they are not generally cast as algorithmic knowledge. Ramanujam [1999] defines an agent to know φ in a model if she can determine that φ is true in the submodel generated by the visible states (the part of the model that the agent sees, such as immediate neighbors in a distributed system), using the model-checking procedure for a standard logic of knowledge. In this case, the knowledge algorithm is simply the model-checking procedure. Another example is recent work on justification logics [Fitting 2005; Artemov and Nogina 2005], based on the intuition that an agent knows φ if she can prove that φ holds in some underlying constructive logic of proofs. The knowledge algorithm in this case consists of searching for a proof of φ.

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.

---

[1] In [Fagin and Halpern 1988], 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.

[2] In [Halpern, Moses, and Vardi 1994], 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.

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 φ 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 $(M, w) \models K_i\varphi$ 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.

Theorem 3.1: [Wansing 1990; Fagin, Halpern, Moses, and Vardi 1995] *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 $\psi \in G$.* [3]

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.)

As we now show, these structures support only propositional reasoning, which we can characterize by the following axiom:

All substitution instances of valid formulas of propositional logic. (*Prop*)

---

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

and the following inference rule:

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

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).*

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 propositionally valid, if we treat formulas of the form $K\varphi$ as primitive propositions. 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.) As we now show, there are constraints on what can be known if we consider KD45 and S5 awareness structures and impossible-worlds structures.

A set of formulas F is *downward closed* if the following conditions hold:

(a) if $\varphi \wedge \psi \in F$, then both φ and ψ 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 (resp., KD45 impossible-worlds structure), $w \in W$, and $w' \in W'$ (resp., $w' \in W \cap W'$). Let $F = \{\varphi \mid (M, w) \models K\varphi\}$ and let $F' = \{\psi \mid (M, w') \models \psi\}$. Then*

(a) F' is propositionally consistent downward-closed set of formulas that contains F;

(b) if M is a KD45 impossible-worlds structure then F is k-compatible with F'.

The next result show that the constraints on F described in Proposition 3.3 are the only constraints on F.

Theorem 3.4: *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.*

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

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

Let 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:

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

The key point here is that, as we shall show, a propositionally consistent set of formulas that is downward closed must be consistent with 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:

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

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

Theorem 3.5:

(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.

Corollary 3.6: *The satisfiability problem for the language $\mathcal{L}^K$ with respect to KD45 awareness structures (resp., KD45 impossible-worlds structures, S5 awareness structures) is NP-complete.*

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 [2005], 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 [1990] (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

φ is then the probability of the set of worlds where φ is true. Clearly, if φ and ψ are logically equivalent, then $\ell(\varphi) = \ell(\psi)$ will be true. However, the agent may not recognize that φ and ψ 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 [Fagin and Halpern 1994]. For example, the fact that an agent may not recognize $\ell(\varphi) = \ell(\psi)$ when φ and ψ 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 φ; if φ and ψ 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 φ, 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 $a_1\ell(\varphi_1) + \cdots + a_n\ell(\varphi_n) \geq c$, where $a_1, \ldots, a_n$ and c are integers. (For ease of exposition, we restrict $\varphi_1, \ldots, \varphi_n$ to be propositional formulas in likelihood formulas; however, the techniques presented here can be extended to deal with formulas that allow arbitrary nesting of ℓ 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', π, μ), where (W, W', π) is KD45 (resp., S5) Kripke structure, and μ 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 φ. We then extend the semantics of $\mathcal{L}^K$ with the following rule for interpreting likelihood formulas:

$(M, w) \models a_1\ell(\varphi_1) + \cdots + a_n\ell(\varphi_n) \geq c$ iff $a_1\mu([\![\varphi_1]\!]_M \cap W') + \cdots + a_n\mu([\![\varphi_n]\!]_M \cap W') \geq c$.

Note that the truth of a likelihood formula at a world does not depend on that world; if a likelihood formula is true at some world of structure M, thenit 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 $a_1x_1 + \cdots + a_kx_k + a_{k+1} \leq b_1y_1 + \cdots + b_my_m + b_{m+1}$, 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 $(2x + 3y \leq 5z) \land (x - y \leq 12z) \Rightarrow (3x + 2y \leq 17z)$ 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:

$$\begin{aligned}&\ell(true) = 1\\&\ell(\neg\varphi) = 1 - \ell(\varphi)\\&\ell(\varphi \wedge \psi) + \ell(\varphi \wedge \neg\psi) = \ell(\varphi).\end{aligned}$$

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 μ 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 μ 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 φ and ψ are equivalent formulas. Using arguments similar to those in Theorem 3.4, 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:

$$K\varphi \Rightarrow \ell(\varphi) > 0. \qquad (KL)$$

Let 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:

$$(K\varphi_1 \wedge \ldots \wedge K\varphi_n) \Rightarrow (\psi_1 \vee \ldots \vee \psi_m)$$

if AX$^P_{Ver} \vdash \varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow (\psi_1 \vee \ldots \vee \psi_m)$
and $\psi_1, \ldots, \psi_m$ are likelihood formulas.
(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:

$$K\varphi \Rightarrow \varphi \text{ if } \varphi \text{ is a likelihood formula.} \qquad (\textit{WVer})$$

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

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 φ 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 both satisfiable in impossible-worlds structures: the former requires that there be an impossible possible world that gets positive probability where both φ 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.[4] 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 $Bound$:

$$\ell(\varphi) \geq 0 \wedge \ell(\varphi) \leq 1. \qquad (\textit{Bound})$$

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 AX^B_{imp} as being the core of probabilistic reasoning in impossible-worlds structures. Let AX^B_{Ver} denote the axiom system $\{Prop, MP, Ineq, Bound, Ver, KL\}$. Let KC^P denote the following extension of KC:

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

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

Theorem 4.2:

---

[4] We remark that Cozic [2005], 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.

(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 with probabilities.*

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.

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.*

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. For example, 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, when trying to consider possible outcomes of an attack on Iraq, the worlds can be taken to represent the outcomes. An agent simply may be unable to contemplate some of the outcomes of an attack, so cannot consider them possible, let alone know that they will happen or not happen. 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 [Modica and Rustichini 1994; Modica and Rustichini 1999; Dekel, Lipman, and Rustichini 1998; Heifetz, Meier, and Schipper 2003] can essentially be understood in terms of awareness structures where awareness is generated by primitive propositions [Halpern 2001; Halpern and Rêgo 2005].

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, there is no proposition p.

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 one discover properties of knowledge in a given situation. As observed in [Fagin, Halpern, Moses, and Vardi 1995], 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 $\mathtt{A}_a$ for determining what the agent is aware of and an algorithm $\mathtt{A}_k$ 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 φ, the algorithmic approach would simply invoke $\mathtt{A}_a$ to check whether the agent is aware of φ; if so, then the agent invokes $\mathtt{A}_k$. For example, if awareness is generated by primitive propositions, then $\mathtt{A}_a$ is the algorithm that, given query φ, checks whether all the primitive propositions in φ are ones the agent is aware of; and we can take $\mathtt{A}_k$ to be the algorithm that does model checking to see if φ is true in every world of W'. (This can be done in time polynomial in W'; see [Fagin, Halpern, Moses, and Vardi 1995].) In impossible-worlds structures, we can interpret lack of awareness of φ as meaning that neither φ 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 φ that the agent is aware of, exactly one of φ 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 [1986] 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 [2006] 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 [1989] 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 [2002], 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 [1983] 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 [Halpern and Pucella 2002]. 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:

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

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.[5] 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 Eva 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 $\mathcal{C}(w') = \{\varphi \mid \mathtt{A}(\varphi) =$ "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 φ. If neither φ nor $\neg\varphi$ is in $\mathcal{C}(w)$, that just means that the algorithm was not able to determine whether φ was true or false. If the algorithm answers "Yes" to both φ 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. If $cont(m) = \varphi$ is the statement that the content of the message m is φ, then in a world where Alice cannot decrypt φ, neither $cont(m) = \varphi$ and $\neg(cont(m) = \varphi)$ 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).

---

[5] 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 [Halpern, Moses, and Vardi 1994].

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 is not using a particular algorithm to compute her knowledge (at least, not one that can be easily described). 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 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 compute whether n is prime. 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 compute whether n is prime. ∎

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 take. While there has been a great deal of discussion (particularly in the philosophy literature) concerning the "metaphysical status" of impossible worlds (cf. [Stalnaker 1996]), the pragmatics of generating impossible worlds has received comparatively little attention. Hintikka [1975] argues that Rantala's [1975] urn models are suitable candidates for impossible worlds. In decision theory, Lipman [1999] 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 π, as the set of worlds satisfying some condition.[6] 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 $\mathbf{true}$ or $\mathbf{false}$), and $\mathcal{C}$ associates with every world in S a set (possibly inconsistent) of propositional formulas. Test T returns $\mathbf{true}$ 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:

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

We can refine the induced impossible-worlds structure by alotting more resources to test T. Intuitively, as an agent performs more introspection, she can recognize more worlds as being impossible. (Manne [2005] 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

---

[6] 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.

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 Φ. 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 $T(w) = \mathbf{true}$ for all worlds w. Of course, after doing some computation, the agent may realize that, say, n_1 is 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 negation. 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 *Permitted(a,b)* or ¬*Permitted(a,b)* for some agent a and action b, and *Permitted* is not allowed to appear in the antecedent of rules [Halpern and Weissman 2003]. 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 $T(w) = \mathbf{true}$ 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 $\mathit{Student}(a) \land \mathit{Female}(a) \Rightarrow \mathit{Permitted}(a, \mathit{Play\text{-}sports})$, 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 $\mathit{Faculty}(a) \land \mathit{Female}(a) \Rightarrow \neg\mathit{Permitted}(a, \mathit{Play\text{-}sports})$, 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 do so here, focussing essentially on expressiveness 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.

As we said, in the full paper, we also consider the pragmatics of logical omniscience. 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. For example, 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. In some domains, there may be a natural interpretation for the awareness function, but it finding a natural impossible-worlds interpretation may be difficult; in other domains, the situation may be just the opposite. Given the increasing understanding of the importance of awareness in game-theoretic applications (see, for example, [Heifetz, Meier, and Schipper 2003; Halpern and Rêgo 2006a; Halpern and Rêgo 2006b]), these pragmatic issues assume more significance, and deserve further exploration.

References

Artemov, S. and E. Nogina (2005). On epistemic logic with justification. In *Proc. 10th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'05)*, pp. 279–294.

Berman, P., J. Garay, and K. J. Perry (1989). Towards optimal distributed consensus. In *Proc. 30th IEEE Symposium on the Foundations of Computer Science (FOCS'89)*, pp. 410–415.

Cozic, M. (2005). Impossible states at work: Logical omniscience and rational choice. In *Proc. First Paris-Amsterdam Meeting of Young Researchers*.

Dekel, E., B. Lipman, and A. Rustichini (1998). Standard state-space models preclude unawareness. *Econometrica 66*, 159–173.

Dolev, D. and A. C. Yao (1983). On the security of public key protocols. *IEEE Transactions on Information Theory 29*(2), 198–208.

Eberle, R. A. (1974). A logic of believing, knowing and inferring. *Synthese 26*, 356–382.

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

Fagin, R. and J. Y. Halpern (1994). Reasoning about knowledge and probability. *Journal of the ACM 41*(2), 340–367.

Fagin, R., J. Y. Halpern, and N. Megiddo (1990). A logic for reasoning about probabilities. *Information and Computation 87*(1/2), 78–128.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. MIT Press.

Fitting, M. (2005). A logic of explicit knowledge. In L. Behounek and M. Bilkova (Eds.), *Logica Yearbook 2004*, pp. 11–22. Filosophia.

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

Halpern, J. Y., Y. Moses, and M. Y. Vardi (1994). Algorithmic knowledge. In *Proc. 5th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'94)*, pp. 255–266. Morgan Kaufmann.

Halpern, J. Y. and R. Pucella (2002). Modeling adversaries in a logic for reasoning about security protocols. In *Proc. Workshop on Formal Aspects of Security (FASec'02)*, Volume 2629 of *Lecture Notes in Computer Science*, pp. 115–132.

Halpern, J. Y. and R. Pucella (2007). Dealing with logical omniscience: Expressiveness and pragmatics. Preprint arXiv:cs.LO/0702011.

Halpern, J. Y. and L. C. Rêgo (2005). Interactive unawareness revisited. In *Theoretical Aspects of Rationality and Knowledge: Proc. Tenth Conference (TARK 2005)*, pp. 78–91.

Halpern, J. Y. and L. C. Rêgo (2006a). Extensive games with possibly unaware players. In *Proc. Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 744–751. Full version available at arxiv.org/abs/0704.2014.

Halpern, J. Y. and L. C. Rêgo (2006b). Reasoning about knowledge of unawareness. In *Principles of Knowledge Representation and Reasoning: Proc. Tenth International Conference (KR '06)*, pp. 6–13. Full version available at arxiv.org/cs.LO/0603020.

Halpern, J. Y. and V. Weissman (2003). Using first-order logic to reason about policies. In *Proc. 16th IEEE Computer Security Foundations Workshop*, pp. 187–201.

Heifetz, A., M. Meier, and B. Schipper (2003). Multi-person unawareness. In *Proc. 9th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'03)*, pp. 148–158. To appear in *Journal of Economic Theory* under the title "Interactive Unawareness".

Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic 4*, 475–484.

Konolige, K. (1986). *A Deduction Model of Belief*. San Francisco: Morgan Kaufmann.

Lipman, B. L. (1999). Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality. *The Review of Economic Studies 66*(2), 339–361.

Manne, K. (2005). Towards a solution to the problem of logical omniscience for resource-bounded agents: An impossible worlds approach to temporal epistemic logic. In *Proc. Conference of the Australasian Association of Philosophy*.

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

Modica, S. and A. Rustichini (1999). Unawareness and partitional information structures. *Games and Economic Behavior 27*(2), 265–298.

Moore, R. C. and G. Hendrix (1979). Computational models of beliefs and the semantics of belief sentences. Technical Note 187, SRI International, Menlo Park, Calif.

Moreno, A. (1998). Avoiding logical omniscience and perfect reasoning: A survey. *AI Communications 11*(2), 101–122.

Moses, Y. and M. R. Tuttle (1988). Programming simultaneous actions using common knowledge. *Algorithmica 3*, 121–169.

Pucella, R. (2006). Deductive algorithmic knowledge. *Journal of Logic and Computation 16*(2), 287–309.

Ramanujam, R. (1999). View-based explicit knowledge. *Annals of Pure and Applied Logic 96*(1–3), 343–368.

Rantala, V. (1975). Urn models: a new kind of non-standard model for first-order logic. *Journal of Philosophical Logic 4*, 455–474.

Rantala, V. (1982). Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica 35*, 18–24.

Stalnaker, R. (1996). Impossibilities. *Philosophical Topics 24*, 193–204.

Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica 49*(4), 523–539.