Belief, Awareness, and Limited Reasoning*

Ronald Fagin and Joseph Y. Halpern
IBM Almaden Research Center, San Jose, CA 95120, U.S.A.

Recommended by Daniel G. Bobrow

ABSTRACT

Several new logics for belief and knowledge are introduced and studied, all of which have the property that agents are not logically omniscient. In particular, in these logics, the set of beliefs of an agent does not necessarily contain all valid formulas. Thus, these logics are more suitable than traditional logics for modelling beliefs of humans (or machines) with limited reasoning capabilities. Our first logic is essentially an extension of Levèsque's logic of implicit and explicit belief, where we extend to allow multiple agents and higher-level belief (i.e., beliefs about beliefs). Our second logic deals explicitly with "awareness," where, roughly speaking, it is necessary to be aware of a concept before one can have beliefs about it. Our third logic gives a model of "local reasoning," where an agent is viewed as a "society of minds," each with its own cluster of beliefs, which may contradict each other.

> The animal knows, of course. But it certainly does not know that it knows.
>
> *Teilhard de Chardin*

1. Introduction

There has long been interest in both philosophy and AI in finding natural semantics for logics of knowledge and belief. The standard approach has been the so-called *possible-worlds* model. The intuitive idea, which goes back to Hintikka [17], is that besides the true state of affairs, there are a number of other possible states of affairs, or possible worlds. Some of these possible worlds may be indistinguishable from the true world to an agent. An agent is then said to *know* or *believe* fact φ if φ is true in all the worlds he thinks possible.

* A preliminary version of this paper appeared in the *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, Los Angeles, CA, 1985. Editor's note: This paper won the best paper award for the conference.

As has been frequently pointed out in the literature (see, for example, [18]), possible-worlds semantics for knowledge and belief do not seem appropriate for modelling human reasoning since they suffer from the problem of what Hintikka calls *logical omniscience*. In particular, this means that agents are assumed to be so intelligent that they must know all valid formulas, and that their knowledge is closed under implication, so that if an agent knows p, and knows that p implies q, then the agent must also know q.

Unfortunately, in real life people are certainly not omniscient. Indeed, possible-worlds advocates have always stressed that this style of semantics assumes an "ideal" rational reasoner, with infinite computational powers. But for many applications, one would like a logic that provides a more realistic representation of human reasoning.

Various attempts to deal with this problem have been proposed in the literature. One approach is essentially syntactic: an agent's beliefs are just described by a set of formulas, not necessarily closed under implication [6, 32], or by the logical consequences of a set of formulas obtained by using an incomplete set of deduction rules [20]. Another approach has been to augment possible worlds by nonclassical "impossible" worlds, where the customary rules of logic do not hold (see, for example, [4, 34, 35]). The syntactic approach lacks the elegance and intuitive appeal of the semantic approach. However, the semantic rules used to assign truth values to the logical connectives in the impossible-worlds approach have tended to be nonintuitive, and it is not clear to what extent this approach has been successful in truly capturing our intuitions about knowledge and belief.

Recently, Levesque [26] has attempted to give an intuitively plausible semantic account of *explicit* and *implicit* belief (where an agent's implicit beliefs include the logical consequences of his explicit belief), essentially by taking partial worlds and a three-valued truth function rather than classical two-valued logic. While we have a number of philosophical and technical criticisms of Levesque's approach (these are detailed in Section 3), it seems to us to be in the right spirit.

Part of the reason that previous semantic attempts to deal with the problem of logical omniscience have failed is that they have not taken into account the fact that it stems from a number of different sources. Among these are:

(1) Lack of awareness: How can someone say that he knows or doesn't know about p if p is a concept he is completely unaware of? One can imagine the puzzled frown on a Bantu tribesman's face when asked if he knows that personal computer prices are going down! The animal (in the quotation at the beginning of the paper) does not know that it knows exactly because it is (presumably) not aware of its knowledge. Similarly, a sentence such as "You're so dumb, you don't even know that you don't know p!" is perhaps best understood as saying "You're not even *aware* that you don't know p".

(2) People are resource-bounded: They lack the computational resources to

deduce all the logical consequences of their knowedge (we still don't know whether Fermat's last theorem is true).

(3) People don't always know the relevant rules: As pointed out by Konolige [20], a student may not know which value of x satisfies the equation $x + a = b$ simply because he doesn't know the rule of subtracting equal quantities from both sides.

(4) People don't focus on all issues simultaneously: Thus, when we say "a believes p," we more properly mean that in a certain frame of mind (when a is focussing on the issues that involve p), it is the case that a believes p. Even if a does perfect reasoning with respect to the limited number of issues on which he is focussing in any given frame of mind, he may not put his conclusions together. Indeed, although in each frame of mind agent a may be consistent, the conclusions a draws in different frames of mind may be inconsistent.

In this paper we present a number of different approaches to modelling lack of logical omniscience. These approaches can be viewed as attempting to model different causes for the lack of omniscience, as suggesed by the discussion above. Our first approach is essentially an extension of Levesque's logic [26] to the multi-agent case, which in addition avoids some of the problems we see in Levesque's approach. This approach is one that attempts to deal with awareness ((1) above). Our second approach combines the possible-worlds framework with a syntactic *awareness* function. The notion of awareness we use in this approach is open to a number of interpretations. One of them is that an agent is aware of a formula if he can compute whether or not it is true in a given situation within a certain time or space bound. This interpretation of awareness gives us a way of capturing resource-bounded reasoning in our model. By adding time into the picture, we can extend the second approach to one that can capture how knowledge is acquired over time, perhaps through the use of a particular (possibly incomplete) set of deduction rules as in [20]. Finally, we present an approach that could be called the *society-of-minds* approach [2, 5, 31], which attempts to capture the type of local reasoning discussed in (4) above (a similar idea has been independently suggested by a number of authors, including Levesque [27], Stalnaker [41], and Zadrozny [44]). The second and third approaches can easily be combined to give a semantics which captures both awareness and local reasoning.

We do not see any of these approaches as being the unique "right" approach to modelling lack of logical omniscience. Rather our philosophy is somewhat more pragmatic. Different notions of knowledge and belief will be appropriate for different applications. We believe that one of the contributions of this paper is providing tools for constructing reasonable semantic models of notions of knowledge with a variety of properties.

The rest of the paper is organized as follows. In the next section we review the "classical" possible-words model of knowledge and belief, to set the stage for our work, while in Section 3 we review Levesque's logic and detail our criticisms of it. In Sections 4, 5, and 6 we describe our three approaches to modelling lack of logical omniscience. In Section 7, we show how we can incorporate time into the picture, allowing us to describe even more situations. In Section 8, we give decision procedures and complete axiomatizations for all the logics we introduce. We conclude in Section 9 with some discussion of our approach and some suggestions for further work.

2. The Possible-Worlds Model

In this section we briefly review possible-worlds semantics for knowledge and belief. The interested reader is encouraged to consult references such as [3, 14] for more details.

Recall that the intuitive idea behind the possible-worlds model is that, besides the true state of affairs, there are a number of other possible states of affairs, or possible worlds. In order to formalize this situation, we first need a language. We stick to propositional logic here, since most of the issues we are interested in dealing with already arise at this level. Besides the standard connectives such as $\wedge$, $\sim$, and $\vee$ from propositional logic, we also need some way to represent belief. We do this by augmenting the language with *modal* operators $L_1, \ldots, L_n$. A formula such as $L_i\varphi$ is read "agent i believes φ."

Formally, we start with a set Φ of primitive propositions, a special formula *true* (this is for convenience only), and close off under negation, conjunction, and the modal operators $L_1, \ldots, L_n$. Thus, if φ and ψ are formulas, then so are $\sim\varphi$, $\varphi \wedge \psi$, and $L_i\varphi$, $i = 1, \ldots, n$. Of course, Boolean connectives such as $\Rightarrow$ and $\vee$ are defined in terms of $\sim$ and $\wedge$ as usual; we take *false* to be an abbreviation of $\sim$*true*. Note we are considering a multi-agent situation here because for many applications we need to reason not only about our own beliefs, but those of other agents. From time to time we consider knowledge rather than belief; in this case we use K_i rather than L_i.

Kripke structures [22] provide a useful formal tool for giving semantics to this language. A Kripke structure M is a tuple $(S, \pi, \mathcal{B}_1, \ldots, \mathcal{B}_n)$, where S is a set of *states* or *possible worlds*, π is an assignment of truth values to the primitive propositions for each state $s \in S$ (so that $\pi(s, p) \in \{\textbf{true}, \textbf{false}\}$ for each state s), and $\mathcal{B}_i$, $i = 1, \ldots, n$, is a binary relation on S which is *serial*, *transitive*, and *Euclidean*. A relation R is *serial* if for each $s \in S$ there is some $t \in S$ such that $(s, t) \in R$; R is *transitive* if $(s, u) \in R$ whenever $(s, t) \in R$ and $(t, u) \in R$; R is *Euclidean* if $(t, u) \in R$ whenever $(s, t) \in R$ and $(s, u) \in R$. Intuitively, $(s, t) \in \mathcal{B}_i$ if in state s, agent i considers state t possible (i.e. if s were the actual state of the world, agent i would consider t a possible state of the world). As we shall see, the conditions on $\mathcal{B}_i$ enforce certain axioms associated with belief. For example the fact that $\mathcal{B}_i$ is serial means that in all worlds, agent i always considers *some* world possible; from this it will follow that he cannot believe in falsehood. By modifying these conditions, we can get different axioms for belief.

We now define a relation $\models$, where $M,s \models \varphi$ is read "φ is *true*, or *satisfied*, in state s of structure M":

$$
\begin{aligned}
&M,s \models \mathit{true},\\
&M,s \models p, \text{ where } p \text{ is a primitive proposition,} \quad \text{iff} \quad \pi(p,s) = \mathbf{true},\\
&M,s \models \sim\varphi \quad \text{iff} \quad M,s \not\models \varphi,\\
&M,s \models \varphi \wedge \psi \quad \text{iff} \quad M,s \models \varphi \text{ and } M,s \models \psi,\\
&M,s \models L_i\varphi \quad \text{iff} \quad M,t \models \varphi \text{ for all } t \text{ such that } (s,t) \in \mathcal{B}_i.
\end{aligned}
$$

The last clause is designed to capture the intuition that agent i believes φ exactly if φ is true in all the worlds that i thinks are possible.[1]

We say a formula φ is *valid in structure M* if $M,s \models \varphi$ for all states s in M; φ is *satisfiable in M* if $M,s \models \varphi$ for some state s in M. We say φ is *valid* if it is valid in all Kripke structures; φ is *satisfiable* if it is satisfiable in some Kripke structure.

This notion of belief can be completely characterized by the following *sound* and *complete* axiom system, traditionally called *weak S5* or *KD45* (cf. [3]). All the axioms given below are valid, the inference rules preserve validity, and every valid formula can be proved from using these axioms and inference rules.

$$\text{All instances of propositional tautologies.} \tag{A1}$$

$$L_i\varphi \wedge L_i(\varphi \Rightarrow \psi) \Rightarrow L_i\psi\ . \tag{A2}$$

$$\sim L_i(\mathit{false})\ . \tag{A3}$$

$$L_i\varphi \Rightarrow L_iL_i\varphi\ . \tag{A4}$$

$$\sim L_i\varphi \Rightarrow L_i \sim L_i\varphi\ . \tag{A5}$$

$$\frac{\varphi,\ \varphi \Rightarrow \psi}{\psi} \quad \text{(modus ponens)}\ . \tag{R1}$$

$$\frac{\varphi}{L_i\varphi}\ . \tag{R2}$$

(A1) and (R1), of course, are holdovers from propositional logic. (A2) says

[1] We remark that for the notion of belief we are considering here, the structure can be simplified if we restrict attention to the one-agent case. Instead of a relation on S, we can simply designate a nonempty subset of S, which we also call $\mathcal{B}$, to be the set of states that the agent considers possible. We can associate with the set $\mathcal{B}$ the binary relation consisting of $\{(s,t) \mid s \in S, t \in \mathcal{B}\}$. It can easily be checked that this relation is serial, Euclidean, and transitive. All the clauses in the definition of $\models$ remain the same, except now the last clause becomes $M,s \models L\varphi$ (we do not need to subscript the L since there is only one agent) iff $M,t \models \varphi$ for all $t \in \mathcal{B}$.

that an agent's beliefs are closed under implication. (A3) says that an agent cannot believe in falsehood. (A4) and (A5) are axioms of introspection. Intuitively, they say that agents are introspective; each agent has complete knowledge about his set of beliefs.

The validity of (A3), (A4), and (A5) is due to the fact that we have taken the $\mathcal{B}_i$ to be serial, transitive, and Euclidean. In a precise sense, (A3) follows from the fact that $\mathcal{B}_i$ is serial, (A4) from the fact that it is transitive, and (A5) from the fact that it is Euclidean. By modifying the properties of the $\mathcal{B}_i$ relations, we can get notions of belief that satisfy different axioms. In particular, the major characteristic taken to distinguish *knowledge* from belief is that if you know something, then it must be true; i.e., $K_i\varphi \Rightarrow \varphi$ holds. This is a stronger statement than (A3). If we take the $\mathcal{B}_i$ relation to be *reflexive* rather than serial, then it turns out that we capture this stronger axiom. (Recall that a relation R on S is reflexive if $(s, s) \in S$ for all $s \in S$, so that a reflexive relation is necessarily serial.)

The classical modal logic of knowledge, called S5, is characterized by the set of axioms above with (A3) replaced by $K_i\varphi \Rightarrow \varphi$ (and all occurrences of L_i replaced by K_i). As suggested above, this can be captured by taking $\mathcal{B}_i$ to be reflexive, transitive, and Euclidean. It is easy to check that a relation is reflexive, transitive, and Euclidean iff it is reflexive, transitive, and symmetric, i.e., an *equivalence relation*. (See [3, 14] for a survey of these issues, as well as a review of the standard techniques of modal logic which give completeness proofs.)

Note that although in KD45 it is possible to have false beliefs (i.e., it is possible that $L_i\varphi$ and $\sim\varphi$ are simultaneously satisfied), it is not possible to believe the negation of a valid formula. Thus, if φ is valid, then we cannot have $L_i\sim\varphi$. This follows directly from our assumption that $\mathcal{B}_i$ is serial. For some applications this is unreasonable (for example, there may be some theorems of mathematics that I believe are false); for such applications, we would drop the assumption that $\mathcal{B}_i$ is serial. We can similarly drop the assumption that $\mathcal{B}_i$ is transitive and Euclidean in cases where axioms (A4) and (A5) are inappropriate. The major advantage of the possible-worlds approach is its flexibility in this regard.

However, the possible-worlds approach seems to commit us to (A2) and (R2). No matter how we modify the $\mathcal{B}_i$ relations, the fact that we say an agent knows or believes a fact exactly if that fact is true in all the worlds the agent considers possible seems to force us to the situation where an agent knows all tautologies and his knowledge is closed under implication. In the remainder of this paper, we show that we can retain the basic intuitions of the possible-worlds approach and still have a logic that avoids the problem of logical omniscience. We present our results in a Kripke-style framework, but we remark that we could have also used the *modal structures* framework of [8, 9].

3. Levesque's Logic of Implicit and Explicit Belief

Before we describe our models for knowledge and belief, we briefly review Levesque's logic of implicit and explicit belief, and discuss our criticisms of it. (We take the liberty of slightly changing Levesque's notation, to make it more consistent with our later development.)

The formulas of the language considered by Levesque are formed in the obvious way, using two modal operators B and L (standing for *explicit belief* and *implicit belief* respectively; an agent's implicit beliefs include all the logical consequences of his explicit beliefs). However, Levesque restricts the language so that no B or L appears within the scope of another. Thus, if φ is a *propositional* formula (does not contain B or L), then $B\varphi$ and $L\varphi$ are also formulas. Levesque does not assume his logic contains the formula *true*.[2]

A *structure for implicit and explicit belief* is a tuple $M = (S, \mathcal{B}, \mathrm{T}, \mathrm{F})$, where S is a set of (primitive) *situations*, $\mathcal{B}$ is a subset of S (the situations that could be the actual ones according to what is believed), and T and F are functions from Φ (the set of primitive propositions) to subsets of S. Intuitively, $\mathrm{T}(p)$ consists of all situations that support the truth of p, while $\mathrm{F}(p)$ consists of all situations that support the falsity of p. We can view this as a modification of the possible-worlds approach (for the one-agent case); instead of possible worlds we have possible situations. It is *not* the case that a primitive proposition is either true or false in a situation; it may be true, false, both, or neither. In particular, we can have a *partial situation* s, that supports neither the truth nor falsity of some primitive proposition p (so that $s \not\in \mathrm{T}(p) \cup \mathrm{F}(p)$) and an *incoherent situation* t that supports both the truth and falsity of some primitive proposition q (so that $t \in \mathrm{T}(q) \cap \mathrm{F}(q)$).

A *complete* situation (called a *possible world* in [26]) is one that supports either the truth or falsity of every primitive proposition and is not incoherent (i.e., s is a member of exactly one of $\mathrm{T}(p)$ and $\mathrm{F}(p)$ for each primitive proposition p). A complete situation s is *compatible* with a situation s' if s and s' agree wherever s' is defined; i.e. if $s' \in \mathrm{T}(p)$ then $s \in \mathrm{T}(p)$, and if $s' \in \mathrm{F}(p)$ then $s \in \mathrm{F}(p)$, for each primitive proposition p. Let $\mathcal{B}^*$ consist of all complete situations in S compatible with some situation in $\mathcal{B}$.

We can now define the *support relations* $\models_\mathrm{T}$ and $\models_\mathrm{F}$ between situations and formulas. Intuitively, $M,s \models_\mathrm{T} \varphi$ when situation s in structure M supports the truth of φ, while $M,s \models_\mathrm{F} \varphi$ when s supports the falsity of φ. The definition is:

$$M,s \models_\mathrm{T} p, \text{ where } p \text{ is a primitive proposition,} \quad \text{iff} \quad s \in \mathrm{T}(p),$$
$$M,s \models_\mathrm{F} p, \text{ where } p \text{ is a primitive proposition,} \quad \text{iff} \quad s \in \mathrm{F}(p);$$

[2] We remark that in the "classical" logic described in the previous section, we could have replaced *true* by $p \vee \sim p$ throughout, where p is any primitive proposition. However, this is not true for Levesque's logic, nor for the logics we present in Sections 4 and 5.

$$\begin{array}{lll} M,s \models_{\mathrm{T}} {\sim}\varphi & \text{iff} & M,s \models_{\mathrm{F}} \varphi\,, \\ M,s \models_{\mathrm{F}} {\sim}\varphi & \text{iff} & M,s \models_{\mathrm{T}} \varphi\,; \end{array}$$

$$\begin{array}{lll} M,s \models_{\mathrm{T}} \varphi_1 \wedge \varphi_2 & \text{iff} & M,s \models_{\mathrm{T}} \varphi_1 \text{ and } M,s \models_{\mathrm{T}} \varphi_2\,, \\ M,s \models_{\mathrm{F}} \varphi_1 \wedge \varphi_2 & \text{iff} & M,s \models_{\mathrm{F}} \varphi_1 \text{ or } M,s \models_{\mathrm{F}} \varphi_2\,; \end{array}$$

$$\begin{array}{lll} M,s \models_{\mathrm{T}} B\varphi & \text{iff} & M,t \models_{\mathrm{T}} \varphi \text{ for all } t \in \mathscr{B}, \\ M,s \models_{\mathrm{F}} B\varphi & \text{iff} & M,s \not\models_{\mathrm{T}} B\varphi\,; \end{array}$$

$$\begin{array}{lll} M,s \models_{\mathrm{T}} L\varphi & \text{iff} & M,t \models_{\mathrm{T}} \varphi \text{ for all } t \in \mathscr{B}^*, \\ M,s \models_{\mathrm{F}} L\varphi & \text{iff} & M,s \not\models_{\mathrm{T}} L\varphi\,. \end{array}$$

We say that the formula φ is *true*, or is *satisfied*, at situation s if $M,s \models_{\mathrm{T}} \varphi$ holds. Levesque defines a formula φ to be valid, written $\models \varphi$, if φ is true at s for all structures $M = (S, \mathscr{B}, \mathrm{T}, \mathrm{F})$, and all *complete* situations $s \in S$.

As Levesque points out, it is easy to see that with this semantics $\models (B\varphi \Rightarrow L\varphi)$, i.e., explicit belief implies implicit belief. It is also easy to see that implicit belief is closed under implication and that all valid propositional formulas are implicitly believed. Thus we have

(1) $\models (L\varphi \wedge L(\varphi \Rightarrow \psi)) \Rightarrow L\psi$, and

(2) if $\models \varphi$ (where φ is propositional), then $\models L\varphi$.

Explicit belief does not seem to suffer from the problems of logical omniscience. Before we go on, let us discuss what we mean by "logical omniscience." An agent is *logically omniscient* if whenever he believes all of the formulas in a set Σ, and Σ logically implies the formula φ, then the agent also believes φ. There are three cases of special interest: (1) what we have been calling *closure under implication* (namely, whenever both φ and $\varphi \Rightarrow \psi$ are believed, then ψ is believed), (2) *closure under valid implication* (if $\varphi \Rightarrow \psi$ is valid, and if φ is believed, then ψ is believed), and (3) belief of valid formulas (if φ is valid, then φ is believed). Explicit belief has none of these three properties. Thus, explicit beliefs are not closed under implication (for example, $Bp \wedge B(p \Rightarrow q) \wedge {\sim}Bq$ is satisfiable), nor under valid implication (although $p \Rightarrow (p \wedge (q \vee {\sim}q))$ is valid, $Bp \wedge {\sim}B(p \wedge (q \vee {\sim}q))$ is satisfiable), and valid formulas are not necessarily believed (${\sim}B(p \vee {\sim}p)$ is satisfiable). Moreover, it is also possible to explicitly believe simultaneously unsatisfiable statements ($Bp \wedge B{\sim}p$ is satisfiable, as, for that matter, is $B(p \wedge {\sim}p)$).

A closer examination of Levesque's semantics shows that the lack of closure under implication and the possibility of believing unsatisfiable statements both stem from the presence of incoherent situations. Indeed, as Levesque points out in [27], while

$$B\varphi \wedge B(\varphi \Rightarrow \psi) \Rightarrow B\psi$$

is not a valid formula, it is easy to check that

$$B\varphi \wedge B(\varphi \Rightarrow \psi) \Rightarrow B(\psi \vee (\varphi \wedge {\sim}\varphi))$$

is valid. Thus, either the agent's beliefs are closed under implication, or else some situation he believes possible is incoherent. (If all the situations that the agent believed possible were coherent, then $\varphi \wedge \sim\varphi$ would not hold in any of them, so ψ would hold in all of them and $B\psi$ would be true.) Similarly, since $B\varphi \wedge B(\sim\varphi) \equiv B(\varphi \wedge \sim\varphi)$, inconsistent beliefs are only possible if every situation the agent believes possible is incoherent (since $\varphi \wedge \sim\varphi$ must be true in every situation the agent believes possible). However, to the extent that $\mathscr{B}$ is viewed as the set of situations that the agent considers possible, it seems unreasonable to allow incoherent situations. It is hard to imagine an agent that woud consider an incoherent situation possible. As Levesque notes in [27], there is a big difference between believing both p and $\sim p$, and believing $p \wedge \sim p$.

On the other hand, in Levesque's logic, an agent's lack of knowledge of valid formulas is not due to incoherent situations, but is rather due to the lack of "awareness" on the part of the agent of some primitive propositions; similar reasons hold for the lack of closure under valid implication. Let us say that an agent is *aware* of a primitive proposition p, which we abbreviate Ap, if $B(p \vee \sim p)$ holds. Thus Ap is true in exactly those situations that support either the truth or falsity of p (they may of course support *both* the truth and falsity of p). Intuitively, this means that p is somehow relevant to the situation and that the agent is "aware" of p in that situation. In the following discussion we use the word "aware" both in the precise mathematical sense just defined, and in the more usual English sense. The reader should be warned, however, that although our mathematical definition does seem to capture some of the properties of the English word, there are several connotations that are certainly *not* captured by the definition. Indeed, in Section 5 we discuss a number of other possible interpretations for the notion of awareness.

Although not every valid formula is believed, it is the case that a valid formula is believed provided that an agent is aware of all the primitive propositions that appear in it. In order to make this precise, given a formula φ, let $\mathrm{Prim}(\varphi)$ be the set of primitive propositions appearing in φ, and let $A\varphi$ be an abbreviation for the conjunction of Ap over all $p \in \mathrm{Prim}(\varphi)$.

Proposition 3.1. *If φ is a valid propositional formula, then $\models A\varphi \Rightarrow B\varphi$.*

Proof. See Appendix.[3] □

Proposition 3.1 suggests that Levesque's semantics may be appropriate for capturing the lack of logical omniscience that arises through lack of awareness,

[3] We remark that the formula $\sim A\varphi \Rightarrow \sim B\varphi$ is *not* valid. For example, if we take φ to be $(p \vee \sim p) \vee (q \vee \sim q)$, then $\sim A\varphi \wedge B\varphi$ is satisfiable in a structure where $\mathscr{B}$ consists of two states, say s and t, such that the agent is aware of only p at s and only q at t.

but not for capturing the type that arises due to lack of computational resources. There may well be a very complicated formula whose truth is hard to figure out, even if you are aware of all the primitive propositions that appear in it.

We have a number of other criticisms, both philosophical and technical, of Levesque's logic:

(1) Although truth (i.e. the $\models_T$ relation) is defined for all situations, only complete situations are considered when checking for validity. This means that there are "valid" formulas φ of Levesque's logic (for example, $p \vee \sim p$) such that $M,s \not\models_T \varphi$ for some situation s. While restricting to complete situations ensures that all propositionally valid formulas continue to be valid in Levesque's logic, it seems inconsistent with the philosophy of looking at situations.

(2) As usual with nonclassical worlds, while the intuitions behind $\models$ seem fairly clear for primitive propositions, they are not so clear for the propositional connectives. For example, suppose that the agent is unaware of the primitive proposition p, so that neither $M,s \models_T p$ nor $M,s \models_F p$ hold. Thus, by the semantic definitions given above, $M,s \models_T (p \equiv p)$ does not hold either. Yet we can still imagine an agent that is unaware of p but is aware of some propositional tautologies, in particular ones like $p \equiv p$. It is interesting to note that in the classical three-valued logic of Lukasiewicz [28], $\Rightarrow$ is usually taken to be a primitive along with $\wedge$ and $\sim$, and the semantics is defined so that $p \equiv p$ is a tautology, even though $p \vee \sim p$ is not. Even though Levesque's semantics could be redefined in this way, the question of motivating the semantics of the connectives still remains.

(3) As Vardi observes [42], although an agent in Levesque's model does not know all the logical consequences of his beliefs (if we understand "logical consequence" to mean "consequence of classical propositional logic"), it follows from Levesque's results [26] that agents in Levesque's logic *are* perfect reasoners in relevance logic [1]. Unfortunately, it seems no more clear that people can do perfect reasoning in relevance logic than that they can do perfect reasoning in classical logic!

Besides the criticisms mentioned above, the current presentation of Levesque's logic suffers from another serious drawback: namely, it deals with only depth-one formulas and with only one agent. But a viable logic of knowledge or belief should be able to capture—within the logic (!)—meta-reasoning about one's own beliefs and reasoning about *other* agents' beliefs. Meta-reasoning is crucial for planning and goal-directed behavior, since one has to reason about the knowledge that one has and needs to acquire. And a knowledge representation utility that does not have certain information may need to reason about where that information is located, and thus about the knowledge of other systems. Such reasoning can quickly get quite complicated, and it is not immediately obvious how to extend Levesque's model to deal with it.

In the next three sections we present three other approaches to dealing with the problem of logical omniscience, each of which attempts to solve aspects of the problem. All of them deal with the multi-agent case and nested beliefs.

4. A Logic of Awareness

The first logic we consider, a logic to reason about awareness, is essentially an extension of Levesque's logic. It allows multiple agents and nested beliefs, both implicit and explicit, while still maintaining many of the properties of Levesque's logic; in particular, it is still the case that explicit belief implies implicit belief. However, we dispense with both partial and incoherent situations. Formally, we proceed as follows. Since we wish to deal with the multi-agent case, we have operators $B_1, \ldots, B_n, L_1, \ldots, L_n$. We allow arbitrary nesting of the B_i and L_j in formulas. A *Kripke structure for awareness* is a tuple $M = (S, \pi, \mathcal{A}_1, \ldots, \mathcal{A}_n, \mathcal{B}_1, \ldots, \mathcal{B}_n)$, where, as in the "classical" possible-worlds model, S is a set of *states*, π is a truth assignment to the primitive propositions for each state $s \in S$, and $\mathcal{B}_i$ is a serial, transitive, Euclidean relation on S for $i = 1, \ldots, n$. We again assume that there is a special formula *true*. The new feature here is $\mathcal{A}_i$ which is a function that associates with each state s a set of primitive propositions. Intuitively, $\mathcal{A}_i(s)$ consists of the primitive propositions of which agent i is aware at state s.

Note that a state corresponds to a complete situation or possible world. There are no partial states. However, we get some of the effects of taking partial states by defining support relations $\models_T^\Psi$ and $\models_F^\Psi$ relative to each set Ψ of primitive propositions. Intuitively, the effect of $\models_T^\Psi$ and $\models_F^\Psi$ is to restrict every state to a partial situation where only the primitive propositions in Ψ are defined. The awareness functions come into play when we consider the semantics of a formula such as $B_i\varphi$. A state s supports the truth of $B_i\varphi$ relative to Ψ if all the states agent i considers possible in s support the truth of φ relative to $\Psi \cap \mathcal{A}_i(s)$, i.e., Ψ further restricted to the set of primitive propositions of which i is aware in state s. We define the set of worlds that agent i considers possible in state s via the $\mathcal{B}_i$ relation, just as in the classical logic of belief. We also define a standard two-valued truth relation $\models$. We define $B_i\varphi$ to be true in state s (i.e., $M,s \models B_i\varphi$) exactly if s supports the truth of $B_i\varphi$ relative to $\mathcal{A}_i(s)$. Implicit belief differs from explicit belief in that for implicit belief we do not take the awareness function into account; all that is relevant is the set of possible worlds.

We now formally define the support relations $\models_T^\Psi$ and $\models_F^\Psi$, and the two-valued notion of truth $\models$:

$M,s \models_T^\Psi$ *true* ,
$M,s \not\models_F^\Psi$ *true* ,
$M,s \models$ *true* ;

$M,s \models_T^\Psi p$, where p is a primitive proposition ,
iff $\pi(s, p) = \mathbf{true}$ and $p \in \Psi$,
$M,s \models_F^\Psi p$, where p is a primitive proposition ,
iff $\pi(s, p) = \mathbf{false}$ and $p \in \Psi$,
$M,s \models p$, where p is a primitive proposition ,
iff $\pi(s, p) = \mathbf{true}$;

$$\begin{array}{lll}
M,s \models_{\mathrm{T}}^{\Psi} {\sim}\varphi & \text{iff} & M,s \models_{\mathrm{F}}^{\Psi} \varphi\ , \\
M,s \models_{\mathrm{F}}^{\Psi} {\sim}\varphi & \text{iff} & M,s \models_{\mathrm{T}}^{\Psi} \varphi\ , \\
M,s \models {\sim}\varphi & \text{iff} & M,s \not\models \varphi\ ;
\end{array}$$

$$\begin{array}{lll}
M,s \models_{\mathrm{T}}^{\Psi} \varphi_1 \wedge \varphi_2 & \text{iff} & M,s \models_{\mathrm{T}}^{\Psi} \varphi_1 \text{ and } M,s \models_{\mathrm{T}}^{\Psi} \varphi_2\ , \\
M,s \models_{\mathrm{F}}^{\Psi} \varphi_1 \wedge \varphi_2 & \text{iff} & M,s \models_{\mathrm{F}}^{\Psi} \varphi_1 \text{ or } M,s \models_{\mathrm{F}}^{\Psi} \varphi_2\ , \\
M,s \models \varphi_1 \wedge \varphi_2 & \text{iff} & M,s \models \varphi_1 \text{ and } M,s \models \varphi_2\ ;
\end{array}$$

$$\begin{array}{lll}
M,s \models_{\mathrm{T}}^{\Psi} B_i\varphi & \text{iff} & M,t \models_{\mathrm{T}}^{\Psi \cap \mathscr{A}_i(s)} \varphi \text{ for all } t \text{ such that } (s,t) \in \mathscr{B}_i\ , \\
M,s \models_{\mathrm{F}}^{\Psi} B_i\varphi & \text{iff} & M,t \models_{\mathrm{F}}^{\Psi \cap \mathscr{A}_i(s)} \varphi \text{ for some } t \text{ such that } (s,t) \in \mathscr{B}_i\ , \\
M,s \models B_i\varphi & \text{iff} & M,s \models_{\mathrm{T}}^{\Phi} B_i\varphi, \text{ where } \Phi \text{ is the set of all primitive} \\
 & & \qquad\qquad \text{propositions}\ ;
\end{array}$$

$$\begin{array}{lll}
M,s \models_{\mathrm{T}}^{\Psi} L_i\varphi & \text{iff} & M,t \models_{\mathrm{T}}^{\Psi} \varphi \text{ for all } t \text{ such that } (s,t) \in \mathscr{B}_i\ , \\
M,s \models_{\mathrm{F}}^{\Psi} L_i\varphi & \text{iff} & M,t \models_{\mathrm{F}}^{\Psi} \varphi \text{ for some } t \text{ such that } (s,t) \in \mathscr{B}_i\ , \\
M,s \models L_i\varphi & \text{iff} & M,t \models \varphi \text{ for all } t \text{ such that } (s,t) \in \mathscr{B}_i\ .
\end{array}$$

Again, we say that φ is valid if $M,s \models \varphi$ for all structures M and all states s in M. We note a number of properties of this definition.

Proposition 4.1.

(1) $\models$ *is complete, i.e., for each* M,s,φ, *either* $M,s \models \varphi$ *or* $M,s \models {\sim}\varphi$.

(2) (a) *If* $\Psi \subseteq \Psi'$ *and if* $M,s \models_{\mathrm{T}}^{\Psi} \varphi$, *then* $M,s \models_{\mathrm{T}}^{\Psi'} \varphi$.

(b) *If* $\Psi \subseteq \Psi'$ *and if* $M,s \models_{\mathrm{F}}^{\Psi} \varphi$, *then* $M,s \models_{\mathrm{F}}^{\Psi'} \varphi$.

(3) (a) *For each set* Ψ *of primitive propositions, if* $M,s \models_{\mathrm{T}}^{\Psi} \varphi$ *then* $M,s \models \varphi$.

(b) *For each set* Ψ *of primitive propositions, if* $M,s \models_{\mathrm{F}}^{\Psi} \varphi$ *then* $M,s \models {\sim}\varphi$.

(4) $\models B_i\varphi \Rightarrow L_i\varphi$.

Proof. Part (1) is immediate from the definition since $M,s \models {\sim}\varphi$ iff $M,s \not\models \varphi$. The proof for parts (2) and (3) proceeds by a straightforward induction on the structure of φ. Part (4) follows easily from 3(a). □

Thus we see that, just as in Levesque's logic, explicit belief implies implicit belief. As we mentioned above, our logic shares a number of other properties with Levesque's. As before, agent i implicitly believes all valid formulas and all the logical consequences of his beliefs. Not all valid formulas are necessarily explicitly believed; in particular, ${\sim}B_i(p \vee {\sim}p)$ is still satisfiable. Neither are an agent's explicit beliefs closed under valid implications; for example, $B_ip \wedge {\sim}B_i(p \wedge (q \vee {\sim}q))$ is satisfiable. Indeed, all of the axioms of Levesque's logic are still sound in our system. (A complete axiomatization for our system is presented in Section 8.) However, because we do not have incoherent situations, our notion of explicit belief differs from Levesque's in that (a) for us, an agent's set of explicit beliefs is closed under implication, and (b) in our system, an agent cannot hold inconsistent beliefs; thus, a formula such as $B_i(p \wedge {\sim}p)$ is not satisfiable.

Unlike Levesque's logic, our logic allows nested beliefs. As expected, (nested) implicit beliefs satisfy all the axioms of weak S5 described in Section 2. We also have $\models (B_i L_i \varphi \equiv B_i \varphi)$, so that agent i explicitly believes that he implicitly believes φ exactly if he explicitly believes φ. Thus, our semantics extends to nested formulas in a reasonable way.

The careful reader will have also noticed one more difference between our logic and Levesque's: namely, the treatment of $\models_{\mathrm{F}}^{\Psi}$ for formulas of the form $B_i\varphi$ and $L_i\varphi$. For Levesque, $M,s \models_{\mathrm{F}} B\varphi$ iff $M,s \not\models_{\mathrm{T}} B\varphi$, so that a situation supports the falsity of explicit belief exactly if it does not support its truth. For us, $M,s \models_{\mathrm{F}}^{\Psi} B_i\varphi$ iff $M,t \models_{\mathrm{F}}^{\Psi \cap \mathscr{A}_i(s)} \varphi$ for some t such that $(s, t) \in \mathscr{B}_i$. Thus, for us, a situation supports the falsity of $B_i\varphi$ iff there is a situation that agent i believes possible that supports the falsity of φ. Essentially this means that the agent has to have positive evidence supporting the falsity of $B_i\varphi$, rather than just no evidence to support the truth of $B_i\varphi$. It turns out that this change has no effect on the valid depth-one formulas (which is why we did not mention it above), but does affect nested formulas. If we had extended Levesque's semantics for $M,s \models_{\mathrm{F}} B\varphi$ in the obvious way, then it would turn out that a formula such as $B_i(B_j p \vee {\sim}B_j p)$ would be valid. Our formulation allows a formula such as ${\sim}B_i(B_j p \vee {\sim}B_j p)$ to be satisfiable (for example, if i is not aware of p). However, since it is not possible for a situation to support the falsity of $p \vee {\sim}p$ in our formulation (although it *can* fail to support its truth), it must be the case that for all sets Ψ of primitive propositions, $M,s \not\models_{\mathrm{F}}^{\Psi} B_i(p \vee {\sim}p)$. Thus it follows that ${\sim}B_j{\sim}B_i(p \vee {\sim}p)$ is a valid formula in our logic of awareness. Intuitively, this says that no agent can have positive evidence that an agent is not aware of p. While certainly this may be an unreasonable property for some applications, it might be quite reasonable for others. We remark that the logic of general awareness presented in the next section does not have this property.[4]

What about the relationship between belief and awareness? Suppose we again define $A_i p$ to be an abbreviation for $B_i(p \vee {\sim}p)$. Note that $M,s \models A_i p$ iff $p \in \mathscr{A}_i(s)$. Again let $A_i\varphi$ be an abbreviation for the conjunction of $A_i p$ taken over all the primitive propositions p that appear in φ. Then it is easy to see that the analogue to Proposition 3.1 holds: if φ is a valid propositional formula, then $\models A_i\varphi \Rightarrow B_i\varphi$. (The proof is straightforward. The key step is to show, by induction on the structure of φ, that if all the primitive propositions in

[4] Lakemeyer [24] has recently presented another extension to Levesque's logic which deals with nested beliefs. He has two types of possibility relations, $\mathscr{B}$ and $\bar{\mathscr{B}}$, to deal with positive beliefs and negative beliefs. Lakemeyer's logic can be extended to deal with multiple agents in such a way that ${\sim}B_i{\sim}B_i(p \vee {\sim}p)$ is not valid.

the propositional formula φ appear in Ψ, then $M,s \not\models \varphi$ iff $M,s \models_{\mathrm{T}}^{\Psi} \varphi$ iff $M,s \not\models_{\mathrm{F}}^{\Psi} \varphi$. The result easily follows.)

There is actually a much deeper relationship between awareness, implicit belief, and explicit belief. For example, it is not hard to show that

$$\begin{aligned}\models B_i(p \vee q)& \\ &\equiv [(A_ip \wedge L_ip) \vee (A_iq \wedge L_iq) \vee (A_ip \wedge A_iq \wedge L_i(p \vee q))]\end{aligned}$$

and that

$$\models B_iB_jp \equiv (A_ip \wedge L_i(A_jp \wedge L_jp)) .$$

Note that in both these cases we were able to capture explicit belief using a combination of implicit belief and awareness. This can always be done. In fact we have:

Proposition 4.2. *Given a formula ψ, we can effectively find a formula ψ^* such that $\models \psi \equiv \psi^*$ and B_i occurs in ψ^* only in the context $B_i(p \vee \sim p)$, where p is a primitive proposition.*

Proof. See Appendix. □

5. A Logic of General Awareness

The logic defined in the previous section limits awareness to primitive propositions. This prevents it from capturing general resource-bounded reasoning. We now present a logic that gives us more fine-grained control over an agent's explicit belief. In particular, in this logic an agent's explicit belief is not closed under implication. The main feature of this logic is an (essentially syntactic) *awareness* operator. Thus, in addition to the modal operators B_i and L_i of the previous logic, we also have a modal operator A_i for each agent i. We can give the formula $A_i\varphi$ a number of interpretations: "i is aware of φ," "i is able to figure out the truth of φ," or even (when reasoning about knowledge bases) "i is able to compute the truth of φ within time T."

A *Kripke structure for general awareness* is a tuple $M = (S, \pi, \mathscr{A}_1, \ldots, \mathscr{A}_n, \mathscr{B}_1, \ldots, \mathscr{B}_n)$, where, as before, S is a set of states, $\pi(s, \cdot)$ is a truth assignment for each state $s \in S$, and $\mathscr{B}_i$ is a serial, transitive, Euclidean relation on S for each agent i.[5] However, we now take $\mathscr{A}_i(s)$ to be an *arbitrary* set of formulas (not just primitive formulas). We do not (yet) place any restrictions on $\mathscr{A}_i(s)$.

[5] Again, to capture knowledge rather than belief, we would take $\mathscr{B}_i$ to be an equivalence relation on S.

In particular, it is possible for both φ and $\sim\varphi$ to be in $\mathscr{A}_i(s)$ (so that the set of formulas an agent is aware of can be inconsistent), for only one of φ and $\sim\varphi$ to be in $\mathscr{A}_i(s)$, or for neither one of φ and $\sim\varphi$ to be in $\mathscr{A}_i(s)$. It is also possible, for example, that $\varphi \wedge \psi$ is in $\mathscr{A}_i(s)$ but $\psi \wedge \varphi$ is not in $\mathscr{A}_i(s)$. The formulas in $\mathscr{A}_i(s)$ are those that the agent is "aware of," not necessarily those he believes.

We have not yet discussed exactly what "awareness" really is, and indeed, we do not intend to do so here at all! The precise interpretation we give to the notion of awareness will depend on the intended application of the logic. By placing various restrictions on the awareness function, we can capture a number of interesting distinct notions. We discuss some interesting restrictions below.

This logic does not have support relations, just a standard two-valued truth relation $\models$, defined inductively as follows:

$$
\begin{aligned}
&M,s \models \mathit{true}\,, \\
&M,s \models p, \text{ where } p \text{ is a primitive proposition, iff } \pi(s,p) = \mathbf{true}\,, \\
&M,s \models \sim\varphi \quad\text{iff}\quad M,s \not\models \varphi\,, \\
&M,s \models \varphi_1 \wedge \varphi_2 \quad\text{iff}\quad M,s \models \varphi_1 \text{ and } M,s \models \varphi_2\,, \\
&M,s \models A_i\varphi \quad\text{iff}\quad \varphi \in \mathscr{A}_i(s)\,, \\
&M,s \models L_i\varphi \quad\text{iff}\quad M,t \models \varphi \text{ for all } t \text{ such that } (s,t) \in \mathscr{B}_i\,, \\
&M,s \models B_i\varphi \quad\text{iff}\quad \varphi \in \mathscr{A}_i(s) \text{ and } M,t \models \varphi \text{ for all } t \text{ such that } (s,t) \in \mathscr{B}_i\,.
\end{aligned}
$$

Note that in this logic, agent i explicitly believes φ iff (1) agent i implicitly believes φ (i.e., φ is true in all the worlds he considers possible) and (2) agent i is aware of φ; thus $B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$. You cannot have explicit beliefs about formulas you are not aware of! If we assume that agents are aware of all formulas, then explicit belief reduces to implicit belief.

It is easy to see that L_i acts like the classical belief operator. Of course, B_i does not. Just as for our previous logic, agents still do not explicitly believe all valid formulas; for example, $\sim B_i(p \vee \sim p)$ is satisfiable because the agent might not be aware of the formula $p \vee \sim p$. However, unlike the previous logic, an agent's explicit beliefs are not necessarily closed under implication; $B_i p \wedge B_i(p \Rightarrow q) \wedge \sim B_i q$ is satisfiable, since i might not be aware of q. Since awareness is essentially a syntactic operator, this approach does suffer from all the shortcomings of the syntactic approach mentioned by Levesque [26]. For example, there is no reason to suppose that $B_i(\varphi \wedge \psi) \equiv B_i(\psi \wedge \varphi)$, since $A_i(\varphi \wedge \psi)$ might hold without $A_i(\psi \wedge \varphi)$ holding. But in fact, people do *not* necessarily identify formulas such as $\psi \wedge \varphi$ and $\varphi \wedge \psi$. Order of presentation does seem to matter. And a computer program that can determine whether $\varphi \wedge \psi$ follows from some initial premises in time τ might not be able to determine whether $\psi \wedge \varphi$ follows from those premises in time τ.

Up to now we have placed no restrictions on the set of formulas that an agent may be aware of. Once we have a concrete interpretation in mind, we may well want to add some restrictions to the awareness function to capture certain properties of "awareness." Because of the clean separation in our framework between belief (captured by the binary relation $\mathcal{B}_i$) and awareness (captured by the syntactic functions $\mathcal{A}_i$), this is quite easy to do. Some typical restrictions we may want to add to $\mathcal{A}_i$ include:

(1) If order of presentation of conjuncts is irrelevant, we could have $\varphi \wedge \psi \in \mathcal{A}_i(s)$ iff $\psi \wedge \varphi \in \mathcal{A}_i(s)$. Similarly, we could decide that an agent is aware of a formula iff he is aware of its negation, so that $\varphi \in \mathcal{A}_i(s)$ iff $\sim\varphi \in \mathcal{A}_i(s)$.

(2) Awareness could be closed under subformulas; i.e., if $\varphi \in \mathcal{A}_i(s)$ and ψ is a subformula of φ, then $\psi \in \mathcal{A}_i(s)$. Note that this makes sense if we are reasoning about a knowledge base that will never compute the truth of a formula unless it has computed the truth of all its subformulas. But it is also easy to imagine a program that knows that $\varphi \vee \sim\varphi$ is true without needing to compute the truth of φ. Perhaps a more reasonable restriction is simply to require that if $\varphi \wedge \psi \in \mathcal{A}_i(s)$ then both φ, $\psi \in \mathcal{A}_i(s)$.[6]

(3) Agent i might only be aware of a certain subset of the primitive propositions, say Ψ. In this case we could take $\mathcal{A}_i(s)$ to consist of exactly those formulas which only mention primitive propositions that appear in Ψ. This type of awareness function gives a logic in somewhat the same spirit as Levesque's logic or the logic of awareness presented in Section 4, but there are some crucial differences. For example, in the awareness logic, the formula $B_i\varphi \Rightarrow B_i(\varphi \vee \psi)$ is valid, whether or not i is aware of ψ; but this formula is not valid in the logic we have just described.

(4) We can allow awareness of agents as well as primitive propositions, so, for example, agent j might not be aware of any formula that mentioned agent i.

(5) A self-reflective agent will be aware of what he is aware of. Semantically, this means that if $\varphi \in \mathcal{A}_i(s)$, then $A_i\varphi \in \mathcal{A}_i(s)$. This corresponds to the axiom $A_i\varphi \Rightarrow A_iA_i\varphi$.

(6) Similarly, an agent might know what formulas he is aware of. Semantically, this means that if $(s, t) \in \mathcal{B}_i$ then $\mathcal{A}_i(s) = \mathcal{A}_i(t)$. This corresponds to the axioms $A_i\varphi \Rightarrow L_iA_i\varphi$ and $\sim A_i\varphi \Rightarrow L_i{\sim}A_i\varphi$. This restriction seems particularly appropriate when awareness is generated by a subset of primitive propositions or a subset of agents, as discussed above.

(7) The elements of $\mathcal{A}_i(s)$ could be exactly those formulas such that agent i can determine in some specified time or space bound whether or not they follow from his information in state s. (See the example at the end of this section for a worked-out example using this notion of awareness.) This type of "awareness" should enable us to provide an abstract model for the notions of

[6] As was pointed out to us by Peter van Emde Boas, without this latter restriction the "pragmatically paradoxical" formula $B_i(p \wedge \sim B_ip)$ ("agent i believes both that p is true and that he doesn't believe it") is satisfiable in the logic (at a state s where $p \wedge \sim B_ip \in \mathcal{A}_i(s)$, but $p \notin \mathcal{A}_i(s)$).

polynomial-time knowledge used in [33]. It might also provide a tool for formalizing the recent advances in cryptography theory. Here the problem is in making sense out of what it means that an adversary does not know how to read a message which is encoded using a public-key cryptosystem (cf. [10, 30, 36]). Such a system is completely insecure from an information-theoretic point of view, but is deemed to be difficult to break in a reasonable amount of time for complexity-theoretic reasons.

We now turn to examining the properties of belief in this logic. First observe that since the semantics of implicit belief (L_i) is identical in the logic of general awareness and in the classical possible-worlds model discussed in Section 2, it is clear that the classical axioms still hold. Indeed, using standard techniques of modal logic, it is easy to show that we can obtain a complete axiomatization of the logic of general awareness simply by adding the axiom $B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$ to the axioms of KD45 discussed in Section 2 (cf. Section 8). However, these axioms do not give us much insight into the properties of explicit belief. In fact, despite the syntactic nature of the awareness operator, explicit belief retains many of the same properties as implicit belief, once we relativize to awareness. For example, corresponding to the axiom $L_i\varphi \wedge L_i(\varphi \Rightarrow \psi) \Rightarrow L_i\psi$ we have:

$$B_i\varphi \wedge B_i(\varphi \Rightarrow \psi) \wedge A_i\psi \Rightarrow B_i\psi \,.$$

Thus, if you explicitly believe φ and $\varphi \Rightarrow \psi$, then you will explicitly believe ψ *provided* you are aware of ψ. Similarly, corresponding to the inference rule that lets us infer $L_i\varphi$ if we have already inferred φ we have:

$$\text{From } \varphi \text{ infer } A_i\varphi \Rightarrow B_i\varphi \,.$$

Again, an agent must be aware of the relevant formula before he explicitly believes it.

Further insight into the relationship between awareness and explicit belief is provided by considering the introspection axioms $L_i\varphi \Rightarrow L_iL_i\varphi$ and $\sim L_i\varphi \Rightarrow L_i\sim L_i\varphi$. Here the most interesting situation arises if we assume $(s, t) \in \mathcal{B}_i$ implies $\mathcal{A}_i(s) = \mathcal{A}_i(t)$, so that an agent knows what formulas he is aware of. In this case, the corresponding properties of explicit belief become:

$$B_i\varphi \wedge A_iB_i\varphi \Rightarrow B_iB_i\varphi$$

and

$$\sim B_i\varphi \wedge A_i(\sim B_i\varphi) \Rightarrow B_i\sim B_i\varphi \,.$$

Again, note that an agent must be aware of the relevant formula before he explicitly believes it. The first of these two axioms shows how, as in the quote from de Chardin, an animal may know, but not know that it knows, while the second indicates how an agent may be "so dumb that he doesn't even know that he doesn't know φ."[7]

[7] Of course, we can construct analogous axioms even if we do not assume that an agent knows what formulas he is aware of, although they are not quite as elegant (cf. [21]).

Naturally, explicit belief will have additional properties once we put some further restrictions on the awareness functions. We now briefly discuss the impact of some of the restrictions discussed above on the properties of explicit knowledge.

The fact that the order of presentation of the conjuncts does not matter can be captured by the axiom $A_i(\varphi \wedge \psi) \equiv A_i(\psi \wedge \varphi)$. It is easy to see that in structures satisfying this restriction $B_i(\varphi \wedge \psi) \equiv B_i(\psi \wedge \varphi)$ is a valid formula. If an agent is aware of a formula iff he is aware of its negation, this can be captured by the axiom $A_i\varphi \equiv A_i{\sim}\varphi$. In structures satisfying this restriction, $B_i\varphi \equiv B_i{\sim}{\sim}\varphi$ is valid.

Taking awareness to be closed under subformulas has some interesting consequences. First note that this property can be captured axiomatically by the axioms schemas

$$A_i({\sim}\varphi) \Rightarrow A_i\varphi\ ,$$

$$A_i(\varphi \wedge \psi) \Rightarrow (A_i\varphi \wedge A_i\psi)\ ,$$

$$A_i(B_j\varphi) \Rightarrow A_i\varphi\ ,$$

$$A_i(L_j\varphi) \Rightarrow A_i\varphi\ ,$$

$$A_i(A_j\varphi) \Rightarrow A_i\varphi.$$[8]

Although agents still do not explicitly believe all valid formulas if awareness is closed under subformulas, it is not hard to show that an agent's beliefs are closed under implication; i.e., $B_i\varphi \wedge B_i(\varphi \Rightarrow \psi) \Rightarrow B_i\psi$ is valid. Thus the seemingly innocuous assumption that awareness is closed under subformulas has rather powerful consequences on the properties of explicit belief. Certainly this assumption is inappropriate for resource-bounded notions of awareness. As we remarked above, it may be easy to see that $\varphi \vee {\sim}\varphi$ is a tautology without having to compute whether either φ or ${\sim}\varphi$ follows from some information. Nevertheless, this observation shows that there are some natural interpretations of awareness and explicit belief (for example, an interpretation of awareness that is closed under subformulas and an interpretation of explicit belief that is not closed under implication) that cannot be simultaneously captured in this framework (cf. [11]). We remark that the model we introduce in the next section overcomes this problem.

[8] We remark that by changing $\Rightarrow$ to $\equiv$ in these axioms, we can capture a notion of awareness generated by a set of primitive propositions (i.e. where $A_i(s)$ consists precisely of the formulas where the only primitive propositions that appear are those in some subset Ψ of primitive propositions).

We close this section with a brief example of an application of this model. Up to now, we have taken an "internal" view of awareness, knowledge, and belief. That is, we have spoken of an agent knowing the formulas he is aware of and having beliefs about his beliefs. This corresponds to the classical view of a reflective agent reasoning about his knowledge and belief. Recently it has been argued that a useful way of analyzing distributed systems and machines is by ascribing knowledge to the processes or components (cf. [13, 37]). The idea here is that we view each process in the system as being in some *local state*; the system as a whole is in a *global state*. Let us use $s(i)$ to denote process i's local state in global state s. Process i is said to *know* φ in global state s if φ is true in all global states s' where $s(i) = s'(i)$. The global states of the system correspond to the possible worlds in a Kripke structure. If we define $\mathcal{B}_i$ so that $(s, s') \in \mathcal{B}_i$ iff $s(i) = s'(i)$, then it is easy to see that this definition makes $\mathcal{B}_i$ into an equivalence relation on the global states. Our definition of knowledge in distributed systems then corresponds precisely to the definition of knowledge in Kripke structures. Thus, this *information-based* notion satisfies all the axioms of S5, the classical logic of knowledge.

We can augment this picture with awareness by assuming that each process is running some algorithm to figure out what it knows. We assume that each process' local state includes some information (perhaps encoded as a set of formulas, although we do not need to assume this). The formulas that i is aware of in global state s (i.e., $\mathcal{A}_i(s)$) are precisely those formulas for which i can determine, using its algorithm, whether or not they follow from the fact that its local state is $s(i)$. In general, i's algorithm will be resource bounded, so that $\mathcal{A}_i(s)$ will not include all formulas. $\mathcal{A}_i(s)$ clearly depends only on $s(i)$, so that if $(s, t) \in \mathcal{B}_i$ (i.e., $s(i) = t(i)$), then we must have $\mathcal{A}_i(s) = \mathcal{A}_i(t)$. Intuitively, we would now like to say that $B_i\varphi$ is true in state s if the algorithm that processor i is running would say that φ is true in state s. However, for this interpretation to be appropriate, we must assume that the algorithm is *correct*: if the algorithm says that φ is true in global state s, then φ must be a consequence of the information contained in the local state $s(i)$, and hence true of all global states where i has the same local state. In particular, $L_i(\varphi)$ must hold. Note we do not need to make any further assumptions on how the algorithm operates. We could imagine that the information in local state $s(i)$ is encoded as a set of formulas and the algorithm applies some deductive procedure to these formulas. Alternatively, we could imagine that the algorithm has information about the set of possible global states and does some "semantic" reasoning about the set of global states t where $s(i) = t(i)$. Using this framework, we now have a way of ascribing explicit as well as implicit knowledge to processes. This might be useful in analyzing systems where we want to view the actions of processes as depending on certain explicit knowledge (cf. the knowledge-based protocols discussed in [12]).

6. A Logic of Local Reasoning

Although the logic of general awareness is quite flexible, it still has the property that an agent cannot hold inconsistent beliefs. In this section we present a logic in which agents can hold inconsistent beliefs that does not make use of incoherent situations.

Our key observation is that one reason that people hold inconsistent beliefs is that beliefs tend to come in non-interacting clusters. We can almost view an agent as a society of minds, each with its own set (or *cluster*) of beliefs, which may contradict each other.

This phenomenon seems to occur even in science. The physicist Eugene Wigner [43] noted that the two great theories physicists reason with are the theory of quantum phenomena and the theory of relativity. However (cf. [35, p. 166]), Wigner thought that the two theories might well be incompatible!

In our previous logics, given a state s, we viewed $\{t \mid (s, t) \in \mathcal{B}_i\}$ as the set of states that agent i thought possible in state s. In our next logic, there is not necessarily one set of states that an agent thinks possible, but rather a number of sets, each one corresponding to a different cluster of beliefs. Alternatively, as discussed in the introduction, we can view these sets as representing the worlds the agent thinks are possible in a given frame of mind, when he is focussing on a certain set of issues.

More formally, a *Kripke structure for local reasoning* is a tuple $M = (S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ where S is a set of *states*, $\pi(s, \cdot)$ is a truth assignment to the primitive propositions for each state $s \in S$ and $\mathcal{C}_i(s)$ is a nonempty set of nonempty subsets of S. If we wish to capture knowledge rather than belief, we need to impose the added condition that s is a member of every set in $\mathcal{C}_i(s)$. Intuitively, if $\mathcal{C}_i(s) = \{T_1, \ldots, T_k\}$, then in state s agent i sometimes (depending perhaps of his state of mind or the issues on which he is focussing) believes that the set of possible states is precisely T_1; sometimes he believes that the set of possible states is precisely T_2, etc. Or we could view each of these sets as representing precisely the worlds that some member of the society in agent i's mind thinks possible. If $\mathcal{C}_i(s)$ is just a singleton set for each state s, say $\{T_s\}$, then this structure is equivalent to the structures of the previous section, where we interpret $(s, t) \in \mathcal{B}_i$ exactly if $t \in T_s$.[9]

The modal operators that seem appropriate for capturing the viewpoint of an agent as a "society of minds" are exactly those discussed in [13] for capturing the knowledge of a group. We now interpret $B_i\varphi$ as "agent i believes φ in *some* frame of mind"; i.e., some member of the society of minds making up i believes φ. Note that although we are using the same symbol in the language,

[9] As we mentioned in the introduction, similar approaches to avoiding logical omniscience were independently discussed by Levesque [27], Stalnaker [41], and Zadrozny [44]. In fact, these models can also be viewed as special cases of minimal models (also known as "neighborhood" or "Scott-Montague" models) that satisfy the added conditions that the set of sets is closed under supersets and is nonempty (cf. [3, Chapter 7]).

this notion is quite different from the notion of explicit belief discussed above. We call this form of explicit belief *local belief*, since it is local to one of the members of the society. We can also imagine a stronger notion where i believes φ in *all* frames of mind or an even stronger notion where it is common knowledge that φ is true in all frames of mind. We could easily add modal operators to the language to describe these notions (and indeed, in an earlier version of this paper, [7], we did have a modal operator to describe the situation where φ was believed in all frames of mind), but in order to be consistent with the operators used in the previous section, we add here only an operator for implicit belief. An agent i *implicitly believes* φ, which we again write $L_i\varphi$, if i would know φ as a result of pooling together the information of his various frames of mind. Although again, this notion has a somewhat different flavor from the notion of implicit belief discussed in previous sections, it does correspond directly to implicit knowledge as defined in [13, 14]; moreover, it is easy to show that explicit (local) belief implies implicit belief. We capture implicit belief formally by saying that agent i implicitly believes φ if φ is true in every world that is considered possible in all frames of mind. By intersecting the set of worlds in this way, we are using the information in each frame of mind to help eliminate possibilities. Of course, if an agent holds inconsistent beliefs in different frames of mind, there will not be any worlds in this intersection, so that he will implicitly believe *false*.

We formally define $\models$ for these structures as follows:

$M,s \models p$, where p is a primitive proposition, iff $\pi(s, p) = \mathbf{true}$,
$M,s \models \sim\varphi$ iff $M,s \not\models \varphi$,
$M,s \models \varphi_1 \wedge \varphi_2$ iff $M,s \models \varphi_1$ and $M,s \models \varphi_2$,
$M,s \models B_i\varphi$ iff there is some $T \in \mathscr{C}_i(s)$ such that $M,t \models \varphi$ for all $t \in T$,
$M,s \models L_i\varphi$ iff $M,t \models \varphi$ for all $t \in \bigcap_{T \in \mathscr{C}_i(s)} T$.

It is easy to see from the semantic definitions given that explicit belief is not closed under implication, but in this case the reason has nothing to do with awareness. The formula $B_ip \wedge B_i(p \Rightarrow q) \wedge \sim B_iq$ is satisfiable simply because in one frame of mind agent i might believe p, in another he might believe $p \Rightarrow q$, but he might never be in a frame of mind where he puts these facts together to conclude q.[10]

More importantly for our purposes, note that an agent may now hold inconsistent beliefs: $B_ip \wedge B_i \sim p$ is satisfiable, since in one frame of mind

[10] We could guarantee closure under implication by requiring that there is always a frame of mind where an agent puts together information that he knows in other frames. Formally this would correspond to the set of sets of possible worlds being closed under intersection, so that if T, $T' \in \mathscr{C}_i(s)$, then $T \cap T' \in \mathscr{C}_i(s)$ (cf. [3, 42]).

agent i might believe p, while in another he might believe $\sim p$. On the other hand, $B_i(p \wedge \sim p)$ is impossible: agents do not believe in incoherent worlds. If we consider knowledge rather than belief (so that s is an element of every member of $\mathscr{C}_i(s)$), then inconsistent beliefs are impossible. Indeed, in this case we get the axiom $B_i\varphi \Rightarrow \varphi$.

Since implicit belief results by pooling together the information available in each frame of mind, we clearly have $B_i\varphi \Rightarrow L_i\varphi$. In particular, it follows that if an agent holds inconsistent beliefs, he implicitly believes everything for vacuous reasons. Thus we get $B_i\varphi \wedge B_i(\sim\varphi) \Rightarrow L_i(\mathit{false})$. (Note that this situation is impossible if we consider knowledge rather than belief, since with knowledge we still have the axiom $B_i\varphi \Rightarrow \varphi$.)

It is easy to see that as defined here, L_i does not satisfy the axioms of KD45, the classical logic of belief. We have just pointed out that $L_i(\mathit{false})$ is consistent, so the axiom (A3) of Section 2 does not hold. (A4) and (A5) do not hold either, although (A1) and (A2) do.

In the classical possible-worlds framework, we can capture various properties of knowledge and belief by imposing various conditions on the binary relation $\mathscr{B}_i$. Analogously, in this framework, we can capture various properties of knowledge and belief by imposing conditions on the set of sets $\mathscr{C}_i(s)$. In the possible-worlds model we have the condition of seriality, which results in the axiom $\sim L_i(\mathit{false})$. The fact that $\mathscr{C}_i(s)$ consists of nonempty sets ensures the validity of $\sim B_i(\mathit{false})$. If we want $\sim L_i(\mathit{false})$ to hold in this logic, we must add the condition that the intersection of the sets in $\mathscr{C}_i(s)$ is nonempty. As remarked above, if we want to capture knowledge rather than belief (so that both $L_i\varphi \Rightarrow \varphi$ and $B_i\varphi \Rightarrow \varphi$ are valid), then we must add the further restriction that s is a member of every member of $\mathscr{C}_i(s)$. It is also not hard to check that if we require that in each frame of mind an agent considers it possible that he is in that frame of mind (that is, if $s' \in T \in \mathscr{C}_i(s)$, then $T \in \mathscr{C}_i(s')$), this ensures the validity of both $B_i\varphi \Rightarrow B_iB_i\varphi$ and $L_i\varphi \Rightarrow L_iL_i\varphi$. Finally, adding the condition that for all $T \in \mathscr{C}_i(s)$ and all $t \in T$ we have $\mathscr{C}_i(t) \subseteq \mathscr{C}_i(s)$, then we have both $\sim B_i\varphi \Rightarrow B_i\sim B_i\varphi$ and $\sim L_i\varphi \Rightarrow L_i\sim L_i\varphi$. (See [3, 42] for a related discussion.)

A particularly interesting special case we can capture is one where in each frame of mind, an agent refuses to admit that he may occasionally be in another frame of mind. (This phenomenon can certainly be observed with people!) Semantically, we can capture this by requiring that if $s' \in T \in \mathscr{C}_i(s)$, then $\mathscr{C}_i(s')$ is the singleton set $\{T\}$.[11] A Kripke structure for local reasoning that satisfies this additional restriction is called a *Kripke structure for narrow-minded agents*.

A narrow-minded agent will believe he is consistent (even if he is not), since

[11] Note that this restriction is not possible in general when dealing with knowledge rather than belief. You cannot refuse to *know* the truth, although you can refuse to *believe* it!

in a given frame of mind he refuses to recognize that he may have other frames of mind. Thus, $B_i(\sim(B_ip \wedge B_i\sim p))$ is valid in this case, even though $B_ip \wedge B_i\sim p$ is consistent. In fact, a stronger statement is true. In *all* frames of mind an agent believes he is consistent. Moreover, since an agent can do perfect reasoning within a given frame of mind, a narrow-minded agent will also believe he is a perfect reasoner. Thus $B_i(B_ip \wedge B_i(p \Rightarrow q) \Rightarrow B_iq)$ is a valid formula in all Kripke structures for narrow-minded agents.

Note that in both the general and the narrow-minded versions of the logic of local reasoning, an agent's beliefs are closed under valid implication (so that if $\varphi \Rightarrow \psi$ is valid, so is $B_i\varphi \Rightarrow B_i\psi$) and agents believe all valid formulas. This is because we have assumed that agents can do perfect reasoning within each cluster. By adding an awareness function to a structure for local reasoning, we can get a model for belief where agents do not necessarily believe all valid formulas. We can then construct a model for a notion of belief and awareness where an agent's awareness is closed under subformulas, but his explicit (local) beliefs are still not closed under logical consequence.

7. Incorporating Time

Even greater flexibility can be attained by incorporating time into the language. Once we can explicitly talk about time, we are in a position to discuss notions like knowledge acquisition and forgetting. Fortunately, it is easy to incorporate time into the possible-worlds framework by adding a relation, and a corresponding modal operator, to capture time. For example, a *Kripke structure for general awareness and time* is a tuple $M = (S, \pi, \mathscr{A}_1, \ldots, \mathscr{A}_n, \mathscr{B}_1, \ldots, \mathscr{B}_n, \mathscr{T})$, where $\mathscr{T}$ is a deterministic, serial relation; i.e. for all $s \in S$, there is a unique $t \in S$ such that $(s, t) \in \mathscr{T}$. Intuitively, $(s, t) \in \mathscr{T}$ if t describes the state of the world at the "next" time instant after s.[12] We also add unary modal operators $\bigcirc$ and $\Diamond$ into the language, where $\bigcirc\varphi$ is true if φ is true at the next time instant (or "tomorrow"), and $\Diamond\varphi$ is true if φ is eventually true. We define $\mathscr{T}^*$ to be the reflexive, transitive closure of $\mathscr{T}$, that is, the binary relation on S defined by $(s, t) \in \mathscr{T}^*$ iff there exist states $s_0, \ldots, s_k$ such that $s = s_0$, $t = s_k$, and $(s_i, s_{i+1}) \in \mathscr{T}$ for $i < k$. More formally, we have:

$$\begin{array}{lll} M,s \models \bigcirc\varphi & \text{iff} & M,t \models \varphi \text{ for (the unique) } t \text{ such that } (s,t) \in \mathscr{T}\,, \\ M,s \models \Diamond\varphi & \text{iff} & M,t \models \varphi \text{ for some } t \text{ such that } (s,t) \in \mathscr{T}^*\,. \end{array}$$

As usual in the literature, we define $\Box$ to be the dual of $\Diamond$, so that $\Box\varphi$ is $\sim\Diamond\sim\varphi$. Thus $\Box\varphi$ is true if φ is true now and forever in the future.

Once we have time in the picture, we can consider investigating what happens when we impose a number of additional constraints on the relation-

[12] Thus we have taken time to be *linear* rather than *branching*, *discrete* rather than *continuous*, and with no endpoint. However, easy modifications can be made to the model presented above to allow us to deal with all of the possibilities (cf. [19]).

ship between belief (or knowledge), time, and awareness. When considering knowledge rather than belief, in some treatments (for example [16, 25, 38]), an additional requirement is placed on the interaction between knowledge and time, which, roughly speaking, captures "not forgetting." The intuition is that the set of worlds an agent thinks possible should decrease over time, as the agent acquires more information. In particular, this means that at a given time, the set of worlds that an agent now thinks could possibly describe the situation of tomorrow is a superset of the set of worlds that he actually thinks possible tomorrow. Syntactically this corresponds to the axiom

$$K_i(\bigcirc\varphi) \Rightarrow \bigcirc K_i\varphi\ ; \tag{1}$$

if agent i knows (today) that φ will be true tomorrow, then tomorrow he will know φ (where we use K_i since we are dealing with knowledge rather than belief). Semantically, this corresponds to the following restriction (where $\mathcal{B}_i$ is an equivalence relation, since we are dealing with knowledge):

> If for some states s, t, u we have $(s, t) \in \mathcal{T}$ and $(t, u) \in \mathcal{B}_i$, then there exists a state w such that $(s, w) \in \mathcal{B}_i$ and $(w, u) \in \mathcal{T}$; i.e. $\mathcal{T} \circ \mathcal{B}_i \subseteq \mathcal{B}_i \circ \mathcal{T}$. (2)

It is easy to check that axiom (1) holds in all structures that obey the restriction (2) (and, as shown in [16], (1) essentially characterizes such structures). As pointed out to us by Elias Thijsse, (1) is not immediately applicable to belief. For example, I may believe now that I may finish writing this paper by tomorrow, but tomorrow I may realize that this belief is false, and no longer believe it. But even with regards to knowledge, (1) is not often not a realistic assumption. People certainly forget! And (2) seems to have rather unpleasant consequences for the decision procedure of the resulting logic (see [16] and Section 8).

Recall that one interpretation we gave the awareness function in Section 5 was in terms of the formulas whose truth could be computed within a certain amount of time. Since we are dealing with a decidable language, we can imagine a program that will *eventually* be able to compute the truth value of every formula. We can capture this very easily in our present framework by simply requiring that the awareness functions satisfy the following constraints:

$$\text{if } (s, t) \in \mathcal{T} \text{ then } \mathcal{A}_i(s) \subseteq \mathcal{A}_i(t) \tag{3}$$

and

> for all $s \in S$ and all formulas φ, there is some t with $(s, t) \in \mathcal{T}^*$ and $\varphi \in \mathcal{A}_i(t)$. (4)

Intuitively, constraint (3) says that agent i's awareness never decreases over time, while (4) says that i is eventually aware of every formula. In a structure satisfying these constraints, we have the following sound inference rule: from φ infer $\Diamond B_i\varphi$. Thus, all valid formulas are eventually believed. Moreover, the obvious weakening of closure under implication also holds. Specifically, as long as $B_i\varphi$ and $B_i(\varphi \Rightarrow \psi)$ are *stable* formulas (once true, they remain true), then if φ and $\varphi \Rightarrow \psi$ are believed, it follows that eventually ψ will be believed too. Thus, if $B_i\varphi$ and $B_i(\varphi \Rightarrow \psi)$ are stable, then we have

$$(B_i\varphi \wedge B_i(\varphi \Rightarrow \psi)) \Rightarrow \Diamond B_i\psi .$$

Other variations on these restrictions are also possible. For example, we may want to drop (4) while retaining (3), so that while an agent's awareness increases, he might not be eventually aware of every formula. We may also want to impose conditions on *how* awareness increases, say by allowing application of a particular deduction rule at every step, where the deduction rule applied might depend on current knowledge or past history (this was suggested to us by Kurt Konolige). There is clearly room for further work here.

If we combine awareness, time, and clusters of belief, we can capture even more complicated situations. For example, it has frequently been observed that people do not like inconsistencies. Yet occasionally they become aware that their beliefs really are inconsistent. When this happens, people tend to modify their beliefs in order to make them consistent again. In a system with local belief, time, and awareness, this can be captured by an axiom such as:

$$(B_i\varphi \wedge B_i{\sim}\varphi) \wedge A_i(B_i\varphi \wedge B_i{\sim}\varphi) \Rightarrow \bigcirc({\sim}(B_i\varphi \wedge B_i{\sim}\varphi)) .$$

This axiom says that if agent i has an inconsistent belief of which he is aware, then at the next state he will modify his belief so that it is not inconsistent.

8. Decision Procedures and Complete Axiomatizations

In the case of the classical logics of belief and knowledge, KD45 and S5, it is known that the problem of deciding whether a formula is satisfiable is NP-complete in the case of one agent, and PSPACE-complete if there is more than one agent (see [14] for a discussion of these results and techniques, many of which go back to Ladner [23]). Despite the apparent extra machinery we have introduced in our models, we can show in most cases that the decision procedures get no harder.

Theorem 8.1.

(1) *The problem of deciding satisfiability of formulas in each of the following logics is NP-complete (and hence the problem of deciding validity is co-NP-complete):*

(a) *Levesque's logic of implicit and explicit belief,*

(b) *the one-agent case of the logic of awareness,*
(c) *the one-agent case of the narrow-minded version of the logic of local reasoning.*

(2) *The problem of deciding satisfiability and validity of formulas is PSPACE-complete in each of the following logics:*
(a) *the multi-agent case of the logic of awareness,*
(b) *the one-agent and multi-agent case of the logic of local reasoning,*
(c) *the multi-agent case of the narrow-minded version of the logic of local reasoning,*
(d) *the one-agent and multi-agent case of each of these logics with time.*

Proof. The proof of these results uses the techniques described in [14, 16], so we only sketch the details here. Since in all of these logics a propositional formula is a tautology iff it is a tautology of propositional logic, the satisfiability problem is NP-hard. The key idea in proving NP-completeness is to show that for each of the logics mentioned in part (1), a satisfiable formula is satisfiable in a small structure: one that has at most polynomially many states (or situations, in the case of Levesque's logic) as the size of the formula. (For all the logics the number of states is in fact linear in the size of the formula except for the narrow-minded version of the logic of local reasoning, where it is quadratic.) We can thus guess a structure for a satisfiable formula in polynomial time, so the problem is in NP.

The PSPACE-completeness results for all the many-knower versions of the logics not involving time follow the same pattern as the PSPACE-completeness result for the many-knower version of S5 discussed in [14]. In particular, the upper bound is proved by showing the existence of a tree-like structure of at most linear depth, while the lower bound involves encoding the operation of a Turing machine that runs in alternating linear time, or alternatively, encoding the satisfiability of QBF formulas (cf. [23]). Because in the logic of local reasoning we have a "society of minds," even the one-knower version of the unrestricted version of this logic has all the necessary features required to get the PSPACE lower bound. Indeed, we could get the lower bound even if we restricted to formulas involving only the L_i operator.

The PSPACE lower bound for the all the logics with time follows from a PSPACE lower bound for the temporal component alone (cf. [15, 40]); the upper bound is proved using techniques similar to those sketched in [16]. □

The logic of general awareness is missing from the list above. Although we conjecture that the one-agent version for this logic is also NP-complete, and the multi-agent version is PSPACE-complete, we have not been able to prove this. The lower bounds still hold, of course, but the best upper bounds we have been able to obtain are nondeterministic exponential time for the one-agent version, and exponential space for the multi-agent version. The difficulty comes from the fact that we have to simultaneously deal with relations of the form $M,s \models_{\mathrm{T}}^{\Psi} \varphi$ for various subsets Ψ of primitive propositions. We remark that we have been able to prove the NP (resp. PSPACE) upper bound for large subclasses of the full language (for example, all those formulas where the outermost occurrences of B_i are in the scope of an even number of negations, or all those formulas where no B_i is in the scope of another B_j). It is also interesting to note that had we dropped the requirement that the $\mathcal{B}_i$ be Euclidean, then the decision procedure would be PSPACE-complete in both the one-agent and multi-agent case.

We also remark that once we add condition (2) as discussed in Section 7 to the semantics of knowledge and time, things can get much worse. As shown in [16], with one knower, the decision procedure becomes complete for double exponential time, while with many knowers it has non-elementary complexity. And with many knowers and the further addition of *common knowledge* (cf. [13, 14]), the logic becomes undecidable (again, see [16]).

We now turn our attention to obtaining complete axiomatizations for all the logics we have discussed. In the logic of awareness described in Section 3, the L_i operator acts exactly like the classical belief operator. Thus, all the axioms for the classical belief operator discussed in Section 2 are still sound in the logic of awareness. In Proposition 4.2 we described a way to effectively transform any formula in the logic into one where the only occurrence of the B_i operator is in the context of $B_i(p \vee \sim p)$ (which we abbreviate $A_i p$). It turns out that no axioms are required to describe $A_i p$, so we get a complete axiomatization for this logic simply by taking the axioms for the classical belief operator and adding the axiom:

$$\varphi \equiv \varphi^* \,, \tag{A6}$$

where φ^* is the formula described in (the proof of) Proposition 4.2. It remains an open question to find more natural axioms that completely characterize the B_i operator.

Theorem 8.2. *The axioms for* KD45 ((A1)–(A5), (R1), (R2)) *together with* (A6) *give a sound and complete axiomatization for the logic of awareness.*

Proof. The fact that (A6) is sound follows from Proposition 4.2. We prove completeness using techniques that go back to Makinson [29], and that are also used to prove completeness of the classical logics of knowledge and belief in [14]. We briefly recall some of the details here.

A formula p is *consistent* (with respect to an axiom system) if $\sim p$ is not provable. A finite set $\{p_1, \ldots, p_k\}$ is consistent exactly if the formula $p_1 \wedge \cdots \wedge p_k$ is consistent. An infinite set of formulas is consistent if every finite subset of it is consistent. A set F of formulas is a *maximal consistent set* if it is consistent and any strict superset is inconsistent. Using standard techniques of propositional reasoning we can show

Lemma 8.3. *In any axiom system that includes* (A1) *and* (R1):

(1) *Any consistent set can be extended to a maximal consistent set.*

(2) *If F is a maximal consistent set, then for all formulas φ and ψ:*

(a) *either $\varphi \in F$ or $\sim\varphi \in F$,*

(b) *$\varphi \wedge \psi \in F$ iff $\varphi \in F$ and $\psi \in F$,*

(c) *if $\varphi \in F$ and $\varphi \Rightarrow \psi \in F$, then $\psi \in F$,*

(d) *if φ is provable, then $\varphi \in F$.* □

In order to prove completeness, we must show that every valid formula is provable. Equivalently, we can prove that every consistent formula is satisfiable. We do so by constructing a *canonical* Kripke structure M^c, containing a state s_V for every maximal consistent set V of formulas, such that $M^c, s_V \models \varphi$ iff $\varphi \in V$. Since every consistent formula is contained in some maximal consistent set, this suffices. Given a set V of formulas, let $V/L_i = \{\varphi \mid L_i\varphi \in V\}$. Let $M^c = (S, \pi, \mathscr{A}_1, \ldots, \mathscr{A}_n, \mathscr{B}_1, \ldots, \mathscr{B}_n)$ be a Kripke structure of awareness where

$$S = \{s_V \mid V \text{ is a maximal consistent set}\},$$

$$\pi(s_V, p) = \begin{cases} \textbf{true}, & \text{if } p \in V, \\ \textbf{false}, & \text{if } p \not\in V, \end{cases}$$

$$\mathscr{A}_i(s_V) = \{p \mid B_i(p \vee \sim p) \in V\},$$

$$\mathscr{B}_i = \{(s_V, s_W) \mid V/L_i \subseteq W\}.$$

As shown in [14], axioms (A3), (A4), and (A5) guarantee that $\mathscr{B}_i$ as defined is indeed serial, transitive, and Euclidean. Now consider the sublanguage $\mathscr{L}'$ consisting of those formulas where the only occurrence of B_i is in the context $B_i(p \vee \sim p)$. Using the techniques of [14], we can easily show by induction on the structure of formulas that for all formulas $\varphi' \in \mathscr{L}'$, we have $M^c, s_V \models \varphi'$ iff $\varphi' \in V$. (The only axioms and rules of inference used in this part of the proof are (A1), (A2), (R1), and (R2).) Finally, suppose $\varphi \in V$. Using (A6), we can find a formula φ^* such that $\varphi \equiv \varphi^*$ is provable and $\varphi^* \in \mathscr{L}'$. By Lemma 8.3, it follows that $\varphi^* \in V$. Since $\varphi^* \in \mathscr{L}'$, we also have $M^c, s_V \models \varphi^*$. Since $\varphi \equiv \varphi^*$ is valid, we have $M^c, s_V \models \varphi$.

We have just shown that if $\varphi \in V$ then $M^c, s_V \models \varphi$. (Actually, from the maximality of V is also easily follows that $\varphi \in V$ iff $M^c, s_V \models \varphi$.) From this we get that if φ is consistent, then for some state s_V in M^c we have $M^c, s_V \models \varphi$. This shows the axiom system is complete. □

In the logic of general awareness, we again have that L_i satisfies all the axioms of KD45. In this logic the explicit belief operator is completely characterized by

$$B_i\varphi \equiv L_i\varphi \wedge A_i\varphi \text{ (explicit belief is equivalent to implicit belief plus awareness)}. \qquad \text{(A7)}$$

Theorem 8.4. *The axioms for* KD45 *together with* (A7) *give a sound and complete axiomatization for the logic of general awareness.*

Proof. Soundness is straightforward, and completeness is proved in a completely analogous fashion to Theorem 8.2. We define the canonical structure in the same way except that now we have $\mathcal{A}_i(s_V) = \{\varphi \mid A_i\varphi \in s_V\}$. Again we can show that $M^c, s_V \models \varphi$ iff $\varphi \in V$. We leave details to the reader. □

Finally, we consider the logic of local reasoning. In this case the L_i operator as defined does *not* satisfy all the axioms of KD45. The only axioms it satisfies are (A1) and (A2), and inference rules (R1) and (R2). (Although, as we remarked above, by imposing extra conditions on $\mathscr{C}_i$, we can obtain axioms (A3), (A4), and (A5).) We also have the following axioms:

$$\sim B_i(\mathit{false}). \qquad \text{(A8)}$$

$$B_i\varphi \Rightarrow L_i\varphi . \qquad \text{(A9)}$$

As we remarked in Section 7, an agent's local beliefs are closed under valid implication and agents believe all valid formulas. This gives us the following rules of inference.

$$\frac{\varphi}{B_i\varphi} . \qquad \text{(R3)}$$

$$\frac{\varphi \Rightarrow \psi}{B_i\varphi \Rightarrow B_i\psi} . \qquad \text{(R4)}$$

Note that (A9) and (R3) render (R2) redundant. Thus we have

Theorem 8.5. *The system consisting of axioms* (A1), (A2), (A8), (A9) *and rules of inference* (R1), (R3), (R4) *is sound and complete for the logic of local reasoning.*

Proof. Again soundness is straightforward. For completeness, we modify the techniques sketched in Theorem 8.2. Again we consider maximal consistent sets of formulas and construct a canonical structure, but in this case the structure has two states corresponding to each maximal consistent set. This technical change allows us to deal with implicit knowledge in a straightforward way.

We define a canonical Kripke structure for local reasoning $M^c = (S, \pi, \mathscr{C}_1, \ldots, \mathscr{C}_n)$ by taking

$$S = \{s^h_V \mid V \text{ is a maximal consistent set, } h = 0, 1\}\,,$$

$$\pi(s^h_V, p) = \begin{cases} \textbf{true}, & \text{if } p \in V, \\ \textbf{false}, & \text{if } p \not\in V, \end{cases} \quad h = 0, 1\,,$$

$$\mathscr{C}_i(s^h_V) = \{T^{h'}_{\psi,V} \mid B_i\psi \in V, h' = 0, 1\}\,,$$

where

$$T^{h'}_{\psi,V} = \{s^{h'}_W \mid \psi \in W\} \cup \{s^l_W \mid V/L_i \subseteq W, l = 0, 1\}\,.$$

In order to show that this is indeed a Kripke structure for local reasoning, we must show that $\mathscr{C}_i(s^h_V)$ is a nonempty set of nonempty subsets of S. Since *true* is provable by (A1), $B_i(\mathit{true})$ is provable by (R3), so we have $B_i(\mathit{true}) \in V$ for all maximal consistent sets V by Lemma 8.3(d). Thus $T^0_{\mathit{true},V}$, $T^1_{\mathit{true},V} \in \mathscr{C}_i(s^h_V)$ and $\mathscr{C}_i(s^h_V)$ is nonempty. To see that $\mathscr{C}_i(s^h_V)$ consists of nonempty sets, suppose $B_i\psi \in V$. Then ψ must be consistent, for if not, $\psi \Rightarrow \mathit{false}$ is provable, and by (R4) and the properties of maximal consistent sets, we would have $B_i(\mathit{false}) \in V$, contradicting the consistency of V by axiom (A8). Since ψ is consistent, there must be some maximal consistent set W containing ψ, so $s^{h'}_W \in T^{h'}_{\psi,V}$.

We next show by induction on structure of formulas that $M^c, s^h_V \models \varphi$ iff $\varphi \in V$. This will show that all consistent formulas are satisfiable, and thus give us completeness of the axiom system. The only interesting cases arise when φ is of the form $B_i\varphi'$ or $L_i\varphi'$.

For $B_i\varphi'$, note that if $B_i\varphi' \in V$, then $T^h_{\varphi',V} \in \mathscr{C}_i(s^h_V)$. By construction, if $s^l_W \in T^h_{\varphi',V}$, then $\varphi' \in W$. (Note that since $B_i\varphi' \in V$, we must have $L_i\varphi' \in V$ by (A9) and the fact that V is a maximal consistent set. Thus if $V/L_i \subseteq W$, then $\varphi' \in W$.) Using the induction hypothesis, it follows that $M^c, t \models \varphi'$ for all $t \in T^h_{\varphi',V}$. Thus $M^c, s^h_V \models B_i\varphi'$. For the converse, suppose $B_i\varphi' \not\in V$. We want to show that $M^c, s^h_V \not\models B_i\varphi'$, so we must show that for all $T^{h'}_{\varphi'',V} \in \mathscr{C}_i(s^h_V)$, there is some $t \in T^{h'}_{\varphi'',V}$ such that $M^c, t \models \sim\varphi'$. But if $T^{h'}_{\varphi'',V} \in \mathscr{C}_i(s^h_V)$, then we must have $B_i\varphi'' \in V$. It must be the case that $\varphi'' \wedge \sim\varphi'$ is consistent. For suppose not. Then $\varphi'' \Rightarrow \varphi'$ is provable, so by (R4) and the fact that V is a maximal consistent set, we must have $B_i\varphi' \in V$, a contradiction. Since $\varphi'' \wedge \sim\varphi'$ is consistent, there must be some maximal consistent set W such that φ'', $\sim\varphi' \in W$. But by construction, $s^h_W \in T^{h'}_{\varphi'',V}$, and by the induction hypothesis we have $M^c, s^h_W \models \sim\varphi'$. Thus s^h_W is the desired state.

For $L_i\varphi'$, note that

$$\bigcap_{T\in\mathscr{C}_i(s_V^h)} T = \{s_W^l \mid V/L_i \subseteq W,\ l=0,1\}\ .$$

(We remark that we took two representatives of each maximal consistent set and defined $\mathscr{C}_i$ the way we did precisely to have this equality hold.) It follows that if $L_i\varphi' \in V$ then $\varphi' \in W$ for all W such that $s_W^l \in \bigcap_{T\in\mathscr{C}_i(s_V^h)} T$. Thus, using the induction hypothesis, $M^c, s_V^h \models L_i\varphi'$. The converse follows along the same lines as the corresponding proof in [14, Theorem 3], so we omit details here. □

We remark that explicit belief in the logic of local reasoning satisfies precisely the axioms of the classical logic EMNP (cf. [3]).

9. Conclusions

We have examined a number of logics, each of which captures different aspects of the problem of lack of logical omniscience, including lack of awareness and local reasoning (within a cluster of beliefs). We expect that other logics can be designed to capture other aspects of this issue.

We view one of the main strengths of our logics to be their semantic naturalness. Thus, the fact that the formula $B_ip \wedge B_i{\sim}p$ is consistent in the logic of local reasoning is not due to some ad hoc condition, but rather follows naturally from the semantic interpretation of B_i. And the relationship between belief and awareness in our logic of general awareness, as exemplified in the axioms, also has a clear semantic interpretation.

We have avoided committing ourselves to a particular notion of awareness in our logic of general awareness, simply because we feel that the conditions on the awareness function should be determined by the particular application. Indeed, we consider the flexibility of this approach to be a point in its favor. Nevertheless, it is clear that further research needs to be done in order to find useful and natural awareness functions. A particularly exciting direction seems to be that of combining awareness and time in interesting ways to try to model interesting properties of knowledge acquisition.

Another interesting direction to take is that of considering quantified versions of these logics. Here some very interesting technical and philosophical questions arise. For example, since we would like to be able to capture sentences such as "He is aware of something that I am not aware of," we seem to be forced into allowing states with different domains, and dealing with all the technical complications that arise there. There is still much work to be done in finding an intuitively motivated logic powerful enough to describe such situations.

Appendix A. Proofs of Proposition 3.1 and Proposition 4.2

Proof of Proposition 3.1. Recall we are trying to show that in Levesque's logic, if φ is a valid propositional formula, then $\models A\varphi \Rightarrow B\varphi$, where $A\varphi$ is the conjunction of Ap taken over all the primitive propositions p that appear in φ, and Ap is an abbreviation for $B(p \vee \sim p)$.

Suppose φ is a valid propositional formula and let $M = (S, \mathcal{B}, \mathrm{T}, \mathrm{F})$ be a structure. We show that in fact $M,s \models_{\mathrm{T}} A\varphi \Rightarrow B\varphi$ holds in *all* situations s in M (and not just in complete situations).

As usual, we define a *valuation* v to be a function that assigns to every primitive proposition a (unique) value in $\{\textbf{true}, \textbf{false}\}$. We can extend a valuation v so that it gives truth values to every propositional formula using the usual rules of propositional logic. Let Ψ be a set of primitive propositions. We define a valuation v to be *compatible with situation s in M with respect to* Ψ if, for all primitive propositions $p \in \Psi$, we have $v(p) = \textbf{true}$ implies $s \in \mathrm{T}(p)$ and $v(p) = \textbf{false}$ implies $s \in \mathrm{F}(p)$. (Note that it may well be that s is an incoherent situation.) Now we can easily show (by induction on the structure of ψ) that if ψ is a propositional formula and v is compatible with s with respect to $\mathrm{Prim}(\psi)$, then $v(\psi) = \textbf{true}$ implies $M,s \models_{\mathrm{T}} \psi$ and $v(\psi) = \textbf{false}$ implies $M,s \models_{\mathrm{F}} \psi$.

We say a set Ψ of primitive propositions is *determined* in a situation s if $M,s \models_{\mathrm{T}} p \vee \sim p$ for each $p \in \Psi$. It easily follows from the definitions that if Ψ is determined in s, then there is *some* valuation v compatible with s with respect to Ψ. Putting together the two observations we have just made, it follows that if ψ is a valid propositional formula and $\mathrm{Prim}(\psi)$ is determined in s, then $M,s \models_{\mathrm{T}} \psi$.

But now returning to our valid formula φ, note that either $\mathrm{Prim}(\varphi)$ is determined in all situations in $\mathcal{B}$ or it isn't. In the former case, by the argument above, φ is true in all situations in $\mathcal{B}$, so $B\varphi$ is true in all situations, as is $A\varphi \Rightarrow B\varphi$. In the latter case, $A\varphi \Rightarrow B\varphi$ holds vacuously in all situations (since $A\varphi$ is false). In either case, $A\varphi \Rightarrow B\varphi$ is true in all situations. □

Proof of Proposition 4.2. For the purposes of this proof, we call a formula *good* if the only occurrences of B_i in the formula are in the context of $B_i(p \vee \sim p)$ (i.e., $A_i p$). Recall that we are trying to prove that for all formulas φ, we can effectively find a good formula φ^* such that $\models \varphi \equiv \varphi^*$. We in fact prove something more general. Given a formula φ and a subformula ψ, we say that an occurrence of ψ in φ appears *positively* in φ if it is in the scope of an even number of negation symbols; otherwise it occurs *negatively*. For example, the first and third p's in the formula $(p \wedge \sim(B_i p \vee B_j \sim p))$ appear positively, while the second occurs negatively.

Lemma A.1. *For all formulas φ, we can effectively find good formulas φ^+, φ^-, and φ^* such that for all structures M, states s, and sets Ψ of primitive propositions, we have*:

$$\begin{array}{lllll}
(1) & M,s \models^{\Psi}_{\mathrm{T}} \varphi & \mathit{iff} & M,s \models^{\Psi}_{\mathrm{T}} \varphi^{+} \ , \\
(2) & M,s \models^{\Psi}_{\mathrm{F}} \varphi & \mathit{iff} & M,s \models^{\Psi}_{\mathrm{F}} \varphi^{-} \ , \\
(3) & M,s \models \varphi & \mathit{iff} & M,s \models \varphi^{*} \ .
\end{array}$$

Moreover, all occurrences of $A_i p$ in φ^+ are positive, while all occurrences of $A_i p$ in φ^- are negative.

Proof. If Ψ is a set of primitive propositions, we define $A_i(\Psi)$ to be an abbreviation for $\bigwedge_{p \in \Psi} A_i p$. We also define $\varphi(\Psi^+)$ (resp. $\varphi(\Psi^-)$) to be the result of replacing all positive occurrences of primitive propositions in φ but *not* in Ψ by *false* (resp. *true*) and negative occurrences of propositions in φ but not in Ψ by *true* (resp. *false*). Thus, if $\Psi = \Phi - \{p\}$ (i.e., Ψ consists of all primitive propositions but p) and we take φ to be $(p \vee q \vee \sim(B_i p \vee B_j {\sim} p))$, then $\varphi(\Psi^+)$ is $(\mathit{false} \vee q \vee \sim(B_i \mathit{true} \vee B_j(\sim \mathit{false})))$. Similarly, if φ is the formula $p \vee \sim p$, then $\varphi(\Psi^+)$ is the formula $\mathit{false} \vee \sim \mathit{true}$, which of course is equivalent to *false*, while $\varphi(\Psi^-)$ is the formula $\mathit{true} \vee \sim \mathit{false}$, which is equivalent to *true*. Note that $\mathrm{Prim}(\varphi(\Psi^+))$ and $\mathrm{Prim}(\varphi(\Psi^-))$ (the primitive propositions that appear in $\varphi(\Psi^+)$ and $\varphi(\Psi^-)$, respectively) are subsets of Ψ.

We now define φ^+, φ^-, and φ^*, by induction on structure:

$$p^+ = p^- = p^* = p$$

for a primitive proposition p;

$$\begin{aligned}
(\sim\varphi)^+ &= \sim(\varphi^-) \ , \\
(\sim\varphi)^- &= \sim(\varphi^+) \ , \\
(\sim\varphi)^* &= \sim(\varphi^*) \ ;
\end{aligned}$$

$$\begin{aligned}
(\varphi \wedge \psi)^+ &= \varphi^+ \wedge \psi^+ \ , \\
(\varphi \wedge \psi)^- &= \varphi^- \wedge \psi^- \ , \\
(\varphi \wedge \psi)^* &= \varphi^* \wedge \psi^* \ ;
\end{aligned}$$

$$\begin{aligned}
(B_i\varphi)^+ &= \bigvee_{\Psi \subseteq \mathrm{Prim}(\varphi)} (A_i(\Psi) \wedge L_i\varphi^+(\Psi^+)) \ , \\
(B_i\varphi)^- &= \bigwedge_{\Psi \subseteq \mathrm{Prim}(\varphi)} (A_i(\Psi) \Rightarrow L_i\varphi^-(\Psi^-)) \ , \\
(B_i\varphi)^* &= (B_i\varphi)^+ \ ;
\end{aligned}$$

$$\begin{aligned}
(L_i\varphi)^+ &= L_i(\varphi^+) \ , \\
(L_i\varphi)^- &= L_i(\varphi^-) \ , \\
(L_i\varphi)^* &= L_i(\varphi^*) \ .
\end{aligned}$$

All the clauses are quite straightforward except that for $B_i\varphi$. The way we replace B_i by L_i here depends on which formulas the agent is aware of. For example, in the definition of $(B_i\varphi)^+$, if he is aware of all the formulas in Ψ, then we replace B_i by L_i, and replace all positive occurrences of primitive propositions not in Ψ by *false*, and all negative occurrences of primitive propositions not in Ψ by *true*. We have taken the disjunction over all possible subsets of Prim(φ) in our translation, since these are the only subsets "relevant" to the formula. We could equally well have taken the disjunction over all subsets of primitive propositions, but then the length of φ^+ would no longer be a function of the length of φ. (We leave it to the reader to check that the length of φ^+, φ^-, and φ^* is at most exponential in the length of φ.)

Not surprisingly, it is straightforward to check that φ^+ and φ^- have all the required properties in all cases except that where φ is of the form $B_i\varphi'$. In order to show that the translation works in this case too, we need a preliminary lemma.

Lemma A.2.

(1) *If* $\Psi' \subseteq \Psi$, *then*
  (a) $M,s \models^{\Psi'}_{\mathrm{T}} \varphi$ *implies* $M,s \models^{\Psi'}_{\mathrm{T}} \varphi(\Psi^+)$, *and*
  (b) $M,s \models^{\Psi'}_{\mathrm{F}} \varphi$ *implies* $M,s \models^{\Psi'}_{\mathrm{F}} \varphi(\Psi^-)$.

(2) *If* $\Psi \subseteq \Psi'$, *then*
  (a) $M,s \models^{\Psi'}_{\mathrm{T}} \varphi(\Psi^+)$ *implies* $M,s \models^{\Psi'}_{\mathrm{T}} \varphi$, *and*
  (b) $M,s \models^{\Psi'}_{\mathrm{F}} \varphi(\Psi^-)$ *implies* $M,s \models^{\Psi'}_{\mathrm{F}} \varphi$.

(3) *If* $\Psi \cap \mathrm{Prim}(\varphi) = \Psi' \cap \mathrm{Prim}(\varphi)$, *then*
  (a) $M,s \models^{\Psi}_{\mathrm{T}} \varphi$ *iff* $M,s \models^{\Psi'}_{\mathrm{T}} \varphi$, *and*
  (b) $M,s \models^{\Psi}_{\mathrm{F}} \varphi$ *iff* $M,s \models^{\Psi'}_{\mathrm{F}} \varphi$.

Proof. Each part of the lemma can be proved by a straightforward induction on the structure of φ. The only nontrivial case, surprisingly enough, is when φ is a primitive proposition. For example, in part (1a), if φ is the primitive proposition p, the proof breaks into two cases. If $p \in \Psi'$, then, since $\Psi' \subseteq \Psi$, we must also have $p \in \Psi$ and $p(\Psi^+) = p$. In this case we clearly have $M,s \models^{\Psi'}_{\mathrm{T}} p$ iff $M,s \models^{\Psi'}_{\mathrm{T}} p(\Psi^+)$. On the other hand, if $p \not\in \Psi'$, then $M,s \not\models^{\Psi'}_{\mathrm{T}} p$. We leave the other cases to the reader. □

Now we show that $M,s \models^{\Psi}_{\mathrm{T}} B_i\varphi$ iff $M,s \models^{\Psi}_{\mathrm{T}} (B_i\varphi)^+$. Let $\Sigma = \Psi \cap \mathcal{A}_i(s) \cap \mathrm{Prim}(\varphi)$. Then we have:

$M,s \models^{\Psi}_{\mathrm{T}} B_i\varphi$

iff $M,s \models^{\Psi \cap \mathrm{Prim}(\varphi)}_{\mathrm{T}} B_i\varphi$ (by Lemma A.2(3))

iff $M,t \models^{\Sigma}_{\mathrm{T}} \varphi^+$ for all t such that $(s, t) \in \mathcal{B}_i$
(by definition)

iff $M,t \models^{\Sigma}_{\mathrm{T}} \varphi^+$ for all t such that $(s, t) \in \mathcal{B}_i$
(by the induction hypothesis)

iff $M,t \models^{\Sigma}_{\mathrm{T}} \varphi^+(\Sigma^+)$ for all t such that $(s, t) \in \mathcal{B}_i$
(by Lemmas A.2(1) and A.2(2))

iff $M,t \models^{\Psi}_{\mathrm{T}} \varphi^{+}(\Sigma^{+})$ for all t such that $(s, t) \in \mathscr{B}_i$
(by Lemma A.2(3))

iff $M,s \models^{\Psi}_{\mathrm{T}} L_i\varphi^{+}(\Sigma^{+})$
(by definition) .

Since we also clearly have $M,s \models^{\Psi}_{\mathrm{T}} A_i(\Sigma)$, it follows that $M,s \models^{\Psi}_{\mathrm{T}} A_i(\Sigma) \wedge L_i\varphi^{+}(\Sigma^{+})$, so that $M,s \models^{\Psi}_{\mathrm{T}} (B_i\varphi)^{+}$.

For the converse, suppose that $M,s \models^{\Psi}_{\mathrm{T}} (B_i\varphi)^{+}$. Thus, for some $\Sigma \subseteq \mathrm{Prim}(\varphi)$, we have $M,s \models^{\Psi}_{\mathrm{T}} A_i(\Sigma) \wedge L_i\varphi^{+}(\Sigma^{+})$. We must also have $\Sigma \subseteq \Psi \cap \mathscr{A}_i(s)$ (otherwise it would not be the case that $M,s \models^{\Psi}_{\mathrm{T}} A_i(\Sigma)$). It now follows that

$M,s \models^{\Psi}_{\mathrm{T}} L_i\varphi^{+}(\Sigma^{+})$

iff $M,t \models^{\Psi}_{\mathrm{T}} \varphi^{+}(\Sigma^{+})$ for all t such that $(s, t) \in \mathscr{B}_i$
(by definition)

iff $M,t \models^{\Psi \cap \mathscr{A}_i(s)}_{\mathrm{T}} \varphi^{+}(\Sigma^{+})$ for all t such that $(s, t) \in \mathscr{B}_i$
(by Lemma A.2(3))

implies $M,t \models^{\Psi \cap \mathscr{A}_i(s)}_{\mathrm{T}} \varphi^{+}$ for all t such that $(s, t) \in \mathscr{B}_i$
(by Lemma A.2(2))

iff $M,t \models^{\Psi \cap \mathscr{A}_i(s)}_{\mathrm{T}} \varphi$ for all t such that $(s, t) \in \mathscr{B}_i$
(by induction hypothesis)

iff $M,s \models^{\Psi}_{\mathrm{T}} B_i\varphi$.

The proof that $(B_i\varphi)^{-}$ has the right properties is similar; we leave details to the reader.

To prove that φ^{*} has the right properties, again all cases are straightforward except when φ is of the form $B_i\varphi'$. To deal with this case, we again first need a lemma.

Lemma A.3. *If φ is a formula where all outermost occurrences of subformulas of the form $B_j\psi$ (i.e., all those that are not in the scope of any B_k) occur positively, then for all states s, we have $M,s \models \varphi$ iff $M,s \models^{\Phi}_{\mathrm{T}} \varphi$.*

Proof. Let $M_i\varphi$ be an abbreviaton for $\sim L_i \sim \varphi$. It is easy to see that using the operator M_i, we can rewrite any formula where all outermost occurrences of subformulas of the form $B_j\psi$ occur positively so that the only occurrences of $\sim$ are either inside the scope of a B_j or occur in front of primitive propositions. Now an easy induction on structure of formulas (treating formulas of the form $B_j\psi$ and $M_j\psi$ as base cases) shows that for all formulas φ in this form, $M,s \models \varphi$ iff $M,s \models^{\Phi}_{\mathrm{T}} \varphi$. □

(We remark that this lemma does not hold for arbitrary formulas. For example, if $\mathscr{A}_i(s) = \emptyset$, then we have $M,s \models \sim B_i p$ but $M,s \not\models^{\Phi}_{\mathrm{T}} \sim B_i p$.)

It now follows immediately that $M,s \models B_i\varphi$ iff $M,s \models^{\Phi}_{\mathrm{T}} B_i\varphi$ (by definition) iff $M,s \models^{\Phi}_{\mathrm{T}} (B_i\varphi)^+$ (by Lemma A.1(1)) iff $M,s \models (B_i\varphi)^+$ (by Lemma A.3, since by construction the only occurrences of subformulas of the form $B_j\psi$ in $(B_i\varphi)^+$ occur positively, and in fact only occur in the context $B_j(p \vee \sim p) = A_j p$) iff $M,s \models (B_i\varphi)^*$ (since $(B_i\varphi)^* = (B_i\varphi)^+$). This completes the proof of Proposition 4.2. □

ACKNOWLEDGMENT

We would like to thank Elias Thijsse for his detailed reading of the paper and numerous pertinent comments and criticisms. Comments and criticisms of Kurt Konolige and Moshe Vardi also forced us to clarify a number of our points. Thanks also to Al Borodin, Daphne Koller, Hector Levesque, Yoram Moses, Peter van Emde Boas, Ed Wimmers, and Lenore Zuck for their helpful remarks.

REFERENCES

1. Anderson, A.R. and Belnap, N.D., *Entailment, the Logic of Relevance and Necessity* (Princeton University Press, Princeton, NJ, 1975).
2. Borgida, A. and Imielinski, T., Decision making in committees—a framework for dealing with inconsistency and non-monotonicity, in: *Proceedings Nonmonotonic Reasoning Workshop* (1984) 21–32.
3. Chellas, B.F., *Modal Logic* (Cambridge University Press, New York, 1980).
4. Cresswell, M.J., *Logics and Languages* (Methuen, London, 1973).
5. Doyle, J., A society of mind, in: *Proceedings IJCAI-83*, Karlsruhe, F.R.G. (1983) 309–314.
6. Eberle, R.A., A logic of believing, knowing and inferring, *Synthese* **26** (1974) 356–382.
7. Fagin, R. and Halpern, J.Y., Belief, awareness, and limited reasoning: Preliminary report, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 491–501.
8. Fagin, R., Halpern, J.Y. and Vardi, M.Y., A model-theoretic analysis of knowledge, in: *Proceedings 25th Annual IEEE Symposium on Foundations of Computer Science* (1984) 268–278.
9. Fagin, R. and Vardi, M.Y., An internal semantics for modal logic, in: *Proceedings 17th Annual ACM Symposium on Theory of Computing* (1985) 305–315.
10. Goldwasser, S., Micali, S. and Rackoff, C., The knowledge complexity of interactive proof-systems, in: *Proceedings 17th Annual ACM Symposium on Theory of Computing* (1985) 291–304.
11. Hadley, R.F., Logical omniscience, AI semantics, and models of belief, Simon Fraser University Tech. Rept. LCCR TR 86-3, Burnaby, BC, 1986.
12. Halpern, J.Y. and Fagin, R., A formal model of knowledge, action, and communication in distributed systems: Preliminary report, in: *Proceedings 4th Annual ACM Symposium on the Principles of Distributed Computing* (1985) 224–236.
13. Halpern, J.Y. and Moses, Y.O., Knowledge and common knowledge in a distributed environment, in: *Proceedings 3rd Annual ACM Conference on Principles of Distributed Computing* (1984) 50–61; second revision: IBM Research Rept. RJ 4421, 1986.
14. Halpern, J.Y. and Moses, Y.O., A guide to the modal logics of knowledge and belief, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 480–490.

15. Halpern, J.Y. and Reif, J.H., The propositional dynamic logic of deterministic, well-structured programs, *Theor. Comput. Sci.* **27** (1983) 127–165.
16. Halpern, J.Y. and Vardi, M.Y., The complexity of reasoning about knowledge and time, in: *Proceedings 18th Annual ACM Symposium on Theory of Computing* (1986) 304–315.
17. Hintikka, J., *Knowledge and Belief* (Cornell University Press, Ithaca, NY, 1962).
18. Hintikka, J., Impossible possible worlds vindicated, *J. Philos. Logic* **4** (1975) 475–484.
19. Hughes, G.E. and Cresswell, M.J., *An Introduction to Modal Logic* (Methuen, London, 1968).
20. Konolige, K., Belief and incompleteness, Artificial Intelligence Note 319, SRI International, Menlo Park, CA, 1984.
21. Konolige, K., What awareness isn't: A sentential view of implicit and explicit belief, in: J.Y. Halpern (Ed.), *Theoretical Aspects of Reasoning about Knowledge*: *Proceedings of the 1986 Conference* (Morgan-Kaufmann, Los Altos, CA, 1986) 241–250.
22. Kripke, S., Semantical analysis of modal logic, *Z. Math. Logik Grundl. Math.* **9** (1963) 67–96.
23. Ladner, R., The computational complexity of provability in systems of modal propositional logic, *SIAM J. Comput.* **6**(3) (1977) 467–480.
24. Lakemeyer, G., Tractable meta-reasoning in propositional logics of belief, unpublished manuscript, 1987.
25. Lehmann, D.J., Knowledge, common knowledge, and related puzzles, in: *Proceedings Third Annual ACM Conference on Principles of Distributed Computing* (1984) 62–67.
26. Levesque, H.J., A logic of implicit and explicit belief, in: *Proceedings AAAI-84*, Austin, TX (1984) 198–202; a revised and expanded version: Lab. Tech. Rept. FLAIR #32, Palo Alto, CA, 1984.
27. H.J. Levesque, Global and local consistency and completeness of beliefs, to appear.
28. J. Lukasiewicz, O logice trojwartosciowej (On three-valued logic), *Ruch Filozoficzny* **5** (1920) 169–171.
29. Makinson, D., On some completeness theorems in modal logic, *Z. Math. Logik Grundl. Math.* **12** (1966) 379–384.
30. Merritt, M.J., Cryptographic protocols, Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, 1983.
31. Minsky, M., Plain talk about neurodevelopmental epistemology, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 1083–1092.
32. Moore, R.C. and Hendrix, G., Computational models of beliefs and the semantics of belief sentences, Technical Note 187, SRI International, Menlo Park, CA, 1979.
33. Moses, Y. and Tuttle, M., Programming simultaneous actions using common knowledge, in: *Proceedings 27th Annual Symposium on Foundations of Computer Science* (1986) 208–221.
34. Rantala, V., Impossible worlds semantics and logical omniscience, *Acta Philos. Fennica* **35** (1982) 106–115.
35. Rescher, N. and Brandom, R., *The Logic of Inconsistency* (Rowman and Littlefield, 1979).
36. Rivest, R., Shamir, A. and Adleman, L., A method for obtaining digital signatures and public-key cryptosystems, *Comm. ACM* **21** (2) (1978) 120–126.
37. Rosenschein, S.J., Formal theories of knowledge in AI and robotics, *New Generation Comput.* **3** (1985) 345–357.
38. Sato, M., A study of Kripke-style methods of some modal logics by Gentzen's sequential method, *Publications Research Inst. Math. Sci. Kyoto Univ.* **13** (2) (1977).
39. Segerberg, K., An essay on classical modal logic, Philosophical Studies, Uppsala, 1972.
40. Sistla, A.P. and Clarke, E.M., The complexity of propositional linear temporal logics, *J. ACM* **32** (3) (1985) 240–251.
41. R. Stalnaker, *Inquiry* (MIT Press, Cambridge, MA, 1985).

42. Vardi, M.Y., On epistemic logic and logical omniscience, in: J.Y. Halpern (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference* (Morgan-Kaufmann, Los Altos, CA, 1986) 293–306.
43. Wigner, E.P., The unreasonable effectiveness of mathematics in the natural sciences, *Comm. Pure Appl. Math.* **13** (1960) 1–14.
44. Zadrozny, W., Explicit and implicit beliefs, a solution of a probem of H. Levesque, Unpublished manuscript, 1985.

Received June 1986; revised version received March 1987