Belief, Awareness, and Limited Reasoning: Preliminary Report

Ronald Fagin
Joseph Y. Halpern

IBM Research Laboratory
San Jose, CA 95193

The animal knows, of course. But it certainly does not know that it knows.

Teilhard de Chardin

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 Levesque'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.

1. Introduction

As has been frequently pointed out in the literature (see, for example, [Hi]), 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_t then the agent must also know q.

Unfortunately, in real life people are certainly not omniscient! Indeed, possible-world 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 ([Eb,MH]), or by the logical consequences of a set of formulas obtained by using an incomplete set of deduction rules ([Ko]). Another approach has been to augment possible worlds by non-classical "impossible" worlds, where the customary rules of logic do not hold (see, for example, [Cr,Ra,RB]). 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 [Lev1] 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 the next section), 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 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 simply lack the computational resources to deduce all the logical consequences of their knowledge (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 [Ko], 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 person *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 suggested by the discussion above. Our first approach is essentially an extension of Levesque's [Levi] 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; it seems to be more appropriate for dealing with resource-bounded reasoning, which has a strongly syntactic component. 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 [Ko]. Finally, we present an approach that could be called the *society-of-minds* approach [Mi,Bl,Do], 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 [Lev2], Stalnaker [St], and Zadrozny [Za]). The second and third approaches can easily be combined to give a semantics which captures both awareness and local reasoning.

2. 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 from a set Φ of primitive propositions, using the standard connectives $\sim$ and $\wedge$, as well as 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 and ψ are formulas, then so are $\sim\varphi$ and $\varphi \wedge \psi$; if φ is *propositional* (does not contain B or L) then $B\varphi$ and $L\varphi$ are also formulas. Of course, Boolean connectives such as $\vee$ and $\Rightarrow$ are defined in terms of $\sim$ and $\wedge$ as usual.

A *model of 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.

An *incoherent* situation s is one that is an element of $\mathrm{T}(p) \cap \mathrm{F}(p)$ for some primitive proposition p. Thus an incoherent situation supports both the truth and falsity of some primitive proposition. A *complete* situation (called a *possible world* in [Lev1]) 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 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* $\models_T$ and $\models_F$ between situations and formulas. Intuitively, $M,s \models_T \varphi$ when situation s in model M supports the truth of φ, while $M,s \models_F \varphi$ when s supports the falsity of φ. The definition is:

$M,s \models_T p$, where p is a primitive proposition, if $s \in \mathrm{T}(p)$.

$M,s \models_F p$, where p is a primitive proposition, if $s \in \mathrm{F}(p)$.

$M,s \models_T \sim\varphi$ if $M,s \models_F \varphi$.

$M,s \models_F \sim\varphi$ if $M,s \models_T \varphi$.

$M,s \models_T \varphi_1 \wedge \varphi_2$ if $M,s \models_T \varphi_1$ and $M,s \models_T \varphi_2$.
$M,s \models_F \varphi_1 \wedge \varphi_2$ if $M,s \models_F \varphi_1$ or $M,s \models_F \varphi_2$.

$M,s \models_T B\varphi$ if $M,t \models_T \varphi$ for all $t \in \mathcal{B}$.
$M,s \models_F B\varphi$ if $M,s \not\models_T B\varphi$.

$M,s \models_T L\varphi$ if $M,t \models_T \varphi$ for all $t \in \mathcal{B}^*$.
$M,s \models_F L\varphi$ if $M,s \not\models_T L\varphi$.

We say that the formula φ is *true*, or is *satisfied*, at situation s if $M,s \models_T \varphi$ holds. Levesque defines a formula φ to be *valid*, written $\models\varphi$, if φ is true at s for all models $M = (S, \mathcal{B}, \mathrm{T}, \mathrm{F})$, 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 formulas are implicitly believed. Thus we have

If $\models\varphi$ (where φ is propositional), then $\models L\varphi$, and
$\models(L\varphi \wedge L(\varphi \Rightarrow \psi)) \Rightarrow L\psi$

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). Now 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 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 [Lev2], 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 knowledge is closed under implication, or else some situation he believes possible is incoherent. 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. However, to the extent that $\mathcal{B}$ 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 would consider an incoherent situation possible. As Levesque notes in [Lev2], there is a big difference between believing both p and $\sim p$, and believing $p \wedge \sim p$.

On the other hand, 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. In some of the following discussion, we shall use the word "aware" both in the precise mathematical sense just defined, and in the more usual English sense, since our mathematical notion of "awareness" does seem to model fairly well the English notion. We shall later mention other possible interpretations for the notion of awareness. Although not every valid sentence is believed, we do have the following:

Proposition 2.1. *Let φ be a valid propositional formula, and let $p_1, \ldots, p_k$ be all the primitive propositions that appear in φ. Then* $\models(A(p_1) \wedge \ldots \wedge A(p_k)) \Rightarrow B(\varphi)$.

Intuitively, Proposition 2.1 says that you believe a valid formula provided that you are aware of all primitive propositions that appear in it. This suggests that Levesque's semantics may be appropriate for capturing the lack of logical omniscience that arises through lack of awareness, 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 non-classical 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* yet 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 [Lu], $\Rightarrow$ is usually taken to be a primitive along with $\wedge$ and $\sim$, and the semantics is defined so that $p{=}p$ 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 [Va], 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 [Lev1] that agents in Levesque's logic *are* perfect reasoners in relevance logic [AB]. 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 are presented in a Kripke-style possible-worlds framework. Kripke-style structures were chosen because of their familiarity to most readers; we could just as well have used the *modal structures* framework of [FHV,FV].

3. A logic of awareness

The first logic we consider, a logic to reason about awareness, is essentially an extension of Levesque's logic (to allow multiple agents and nested beliefs) that dispenses with both partial and incoherent situations. Formally, we proceed as follows. Since we wish to deal with many agents, we fix a set of agents (or *players*) $1,\ldots,n$ and instead of a single B and L, we allow operators $B_1,\ldots,B_n,L_1,\ldots,L_n$. We allow arbitrary nesting of the B_i's and L_j's in formulas. A *Kripke structure for awareness* is a tuple $M=(S,\pi,\mathscr{A}_1,\ldots,\mathscr{A}_n,\mathscr{B}_1,\ldots,\mathscr{B}_n)$, where S is a set of *states*, $\pi(s,\cdot)$ is a truth assignment to the primitive propositions for each state $s \in S$ (i.e., $\pi(s,p) \in \{\text{true},\text{false}\}$ for each $p \in \Phi$), $\mathscr{A}_i$ associates with each state s a set of primitive propositions (intuitively, these are the primitive propositions of which player i is aware at state s), and $\mathscr{B}_i$ is a binary relation over S which is transitive, Euclidean, and serial, for each player i.[1] For convenience, we also assume that *false* is a special primitive proposition, and that $\pi(s,\mathit{false})=\text{false}$ and $\mathit{false} \in \mathscr{A}_i(s)$ for all states s.

Note that a state corresponds to a complete situation or possible world. There are no partial states. However, as we shall see, the awareness functions make each state partial from the point of view of player i. Of course, in a given state, player i and player j will not necessarily be aware of the same formulas.

In order to define truth of formulas in this logic with many players, we have support relations $\models_T^{\Psi}$ and $\models_F^{\Psi}$ for 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. We also have a standard two-valued notion of truth defined via $\models$. We proceed as follows:

[1] A relation 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 $(s,u) \in R$; R is *serial* if for each $s \in S$ there is some $t \in S$ such that $(s,t) \in R$. Intuitively, $(s,t) \in \mathscr{B}_i$ if player i in state s believes that state t is possible. It is well known (see [Ch,HM] for discussion and motivation) that by making the $\mathscr{B}_i$ relations transitive, Euclidean, and serial we capture the axioms associated with belief. In particular, the fact that $\mathscr{B}_i$ is transitive ensures the soundness of the axiom $L_i\varphi \Rightarrow L_iL_i\varphi$, the fact that it is Euclidean ensures the soundness of $\sim L_i\varphi \Rightarrow L_i\sim L_i\varphi$, while the fact that it is serial ensures the soundness of $\sim L_i(\mathit{false})$. Note that the fact that $\mathscr{B}_i$ is Euclidean means that for any given state s, the relation $\mathscr{B}_i$ restricted to the worlds possible relative to s (i.e., $\{t \mid (s,t) \in \mathscr{B}_i\}$) is an equivalence relation (i.e., it is reflexive, symmetric, and transitive); the fact that $\mathscr{B}_i$ is serial means that there is always *some* world possible relative to s. Knowledge differs from belief in that you cannot *know* false facts (although you may believe them). This amounts to requiring the additional axiom $K_i\varphi \Rightarrow \varphi$ (note we use K_i to denote "player i knows". In order to capture knowledge, instead of assuming that each $\mathscr{B}_i$ is transitive, Euclidean, and serial, we would make the stronger assumption that each $\mathscr{B}_i$ is an equivalence relation on S.

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

$M,s \models_T^{\Psi} \sim\varphi$ if $M,s \models_F^{\Psi} \varphi$.
$M,s \models_F^{\Psi} \sim\varphi$ if $M,s \models_T^{\Psi} \varphi$.
$M,s \models \sim\varphi$ if $M,s \not\models \varphi$.

$M,s \models_T^{\Psi} \varphi_1 \wedge \varphi_2$ if $M,s \models_T^{\Psi} \varphi_1$ and $M,s \models_T^{\Psi} \varphi_2$.
$M,s \models_F^{\Psi} \varphi_1 \wedge \varphi_2$ if $M,s \models_F^{\Psi} \varphi_1$ or $M,s \models_F^{\Psi} \varphi_2$.
$M,s \models \varphi_1 \wedge \varphi_2$ if $M,s \models \varphi_1$ and $M,s \models \varphi_2$.

$M,s \models_T^{\Psi} B_i\varphi$ if $M,t \models_T^{\Psi \cap \mathcal{A}_i(s)} \varphi$ for all t such that $(s,t) \in \mathcal{B}_i$.
$M,s \models_F^{\Psi} B_i\varphi$ if $M,t \models_F^{\Psi \cap \mathcal{A}_i(s)} \varphi$ for some t such that $(s,t) \in \mathcal{B}_i$.
$M,s \models B_i\varphi$ if $M,s \models_T^{\Phi} B_i\varphi$, where Φ is the set of all primitive propositions.

$M,s \models_T^{\Psi} L_i\varphi$ if $M,t \models_T^{\Psi} \varphi$ for all t such that $(s,t) \in \mathcal{B}_i$.
$M,s \models_F^{\Psi} L_i\varphi$ if $M,t \models_F^{\Psi} \varphi$ for some t such that $(s,t) \in \mathcal{B}_i$.
$M,s \models L_i\varphi$ if $M,t \models \varphi$ for all t such that $(s,t) \in \mathcal{B}_i$.

We note a number of properties of this definition.

Proposition 3.1.

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 $M,s \models_T^{\Psi} \varphi$, then $M,s \models_T^{\Psi'} \varphi$.*
   b. *If $\Psi \subseteq \Psi'$ and if $M,s \models_F^{\Psi} \varphi$, then $M,s \models_F^{\Psi'} \varphi$.*
3. a. *For each set Ψ of primitive propositions, if $M,s \models_T^{\Psi} \varphi$ then $M,s \models \varphi$.*
   b. *For each set Ψ of primitive propositions, if $M,s \models_F^{\Psi} \varphi$ then $M,s \models \sim\varphi$.*

Proof. The proof in each case is a straightforward induction on the structure of φ. ■

From the definitions, it also follows that $B_i\varphi$ is true relative to Ψ at state s (i.e., $M,s \models_T^{\Psi} B_i\varphi$) iff φ is true relative to $\Psi \cap \mathcal{A}_i(s)$ at all the states that player i thinks possible (in state s). In particular, this means that $B_i\varphi$ is true at state s (i.e., $M,s \models B_i\varphi$, that is, $M,s \models_T^{\Phi} B_i\varphi$, where Φ is the set of all primitive propositions) iff φ is true relative to $\mathcal{A}_i(s)$ (the primitive propositions of which i is aware at state s) at all the states player i thinks possible. By way of contrast, since $L_i\varphi$ depends only on player i's implicit belief, and not his awareness, $L_i\varphi$ is true at state s if in all states that player i thinks possible, φ is true (irrespective of whether i is aware of φ). It can easily be shown (using parts 2(a) and 3(a) of Proposition 3.1) that just as in Levesque's logic, we have $\models(B_i\varphi \Rightarrow L_i\varphi)$: if player i explicitly believes φ, then he also implicitly believes φ. Note that we also have $\models(B_iL_i\varphi \equiv B_i\varphi)$, so that player 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.

Our logic of awareness shares a number of properties with Levesque's. As before, player 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 a player's explicit beliefs closed under valid implications; for example, $B_ip \wedge \sim B_i(p \wedge (q \vee \sim q))$ is satisfiable. And Proposition 2.1 still holds, indicating that we also have a logic of awareness. Indeed, all of the axioms of Levesque's logic are still sound in our system. (A complete axiomatization of our logic will appear in the full paper.) 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.

The careful reader will have also noticed one more difference between our logic and Levesque's: namely, the treatment of $\models_F^{\Psi}$ for formulas of the form $B_i\varphi$ and $L_i\varphi$. For Levesque, $M,s \models_F B\varphi$ iff $M,s \not\models_T B\varphi$, so that a situation supports the falsity of explicit belief exactly if it does not support its truth. For us, $M,s \models_F^{\Psi} B_i\varphi$ if $M,t \models_F^{\Psi \cap \mathcal{A}_i(s)} \varphi$ for some t such that $(s,t) \in \mathcal{B}_i$. Thus, for us, a situation supports the falsity of $B_i\varphi$ exactly if there is a situation that agent i believes possible that supports the falsity of φ. 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. Our formulation allows a formula such as $\sim B_i(B_jp \vee \sim B_jp)$ to be satisfiable (for example, if i is not aware of p).

We close this section with one final observation on the relationship between implicit and explicit belief in our logic. It is easy to check that

$$B_i(p \vee q) \equiv$$
$$[(A_ip \wedge L_ip) \vee (A_iq \wedge L_iq) \vee (A_ip \wedge A_iq \wedge L_i(p \vee q))].$$

Similarly, we can show that, for example, $B_iB_jp \equiv (A_ip \wedge L_i(A_jp \wedge L_jp))$. Note that in both these cases explicit belief was replaced by a combination of implicit belief and awareness (recall that A_ip is an abbreviation for $B_i(p \vee \sim p)$. This can be done in general. In fact we have the following proposition, whose proof is given in the full paper:

Proposition 3.2. *Given a formula φ, we can effectively find a formula φ' such that $\varphi \equiv \varphi'$ and B_i occurs in φ' only in the context $B_i(p \vee \sim p)$, where p is a primitive proposition.*

We remark that φ' is in general exponential in the size of φ, so that such a representation may not be very succinct. However, this result does show that in a strong sense explicit belief in this logic is generated by awareness of primitive propositions.

4. 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 a player's awareness. In particular, in this logic, an agent's knowledge is not closed under implication. The main new feature of this logic is a somewhat 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 player 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 transitive, Euclidean, and serial binary relation on S for each player i.[2] However, now we take $\mathscr{A}_i(s)$ to be an arbitrary set of formulas (not just primitive formulas), where again we add the restrictions that $false \in \mathscr{A}_i(s)$ for all s. Note that we have not (yet) placed any restrictions on $\mathscr{A}_i(s)$. In particular, it is possible for both φ and $\sim\varphi$ to be in $\mathscr{A}_i(s)$, and it is possible that, for example, $\varphi \wedge \psi$ 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 at all here! 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 shall discuss some interesting restrictions below.

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

$M,s \models p$, where p is a primitive proposition, if $\pi(s,p) = $ true.

$M,s \models \sim\varphi$ if $M,s \not\models \varphi$.

$M,s \models \varphi_1 \wedge \varphi_2$ if $M,s \models \varphi_1$ and $M,s \models \varphi_2$.

$M,s \models A_i\varphi$ if $\varphi \in \mathscr{A}_i(s)$.

$M,s \models B_i\varphi$ if $\varphi \in \mathscr{A}_i(s)$ and $M,t \models \varphi$ for all t such that $(s,t) \in \mathscr{B}_i$.

$M,s \models L_i\varphi$ if $M,t \models \varphi$ for all t such that $(s,t) \in \mathscr{B}_i$.

Note that with this logic, player i explicitly believes φ iff (a) player i implicitly believes φ (i.e., φ is true in all the worlds he considers possible) and (b) player 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! It is easy to see that L_i acts like the classical belief operator; if we assume that players are aware of all formulas, this logic reduces to the classical logic of belief, known as KD45 or weak S5 (cf. [Ch,HM]).

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; $B_ip \wedge B_i(p \Rightarrow q) \wedge \sim B_iq$ 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 [Lev1]. 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 compute the truth of $\varphi \wedge \psi$ in time T might not be able to compute the truth of $\psi \wedge \varphi$ in time T.

On the other hand, as mentioned above, depending on the intended application, we may want to add some restrictions to the awareness function to capture certain properties of "awareness". These include:

1. Awareness could be closed under subformulas; i.e., if $\varphi \in \mathscr{A}_i(s)$ and ψ is a subformula of φ, then $\psi \in \mathscr{A}_i(s)$. 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 $\varphi \wedge \psi \in \mathscr{A}_i(s)$ then both $\varphi, \psi \in \mathscr{A}_i(s)$.[3]
2. If order of presentation of conjuncts is irrelevant, we could have $\varphi \wedge \psi \in \mathscr{A}_i(s)$ iff $\psi \wedge \varphi \in \mathscr{A}_i(s)$. Similarly, we could decide that an agent is aware

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

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

of a formula iff he is aware of its negation, so that $\varphi \in \mathcal{A}_i(s)$ iff $\sim\varphi \in \mathcal{A}_i(s)$.

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 3, 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 players as well as formulas, so, for example, player j might not be aware of any formula that mentioned player 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 axiom $A_i\varphi \Rightarrow L_iA_i\varphi$. This restriction seems particularly appropriate when awareness is generated by a subset of primitive propositions or a subset of players, as discussed above.
7. The elements of $\mathcal{A}_i(s)$ could be exactly those formulas whose truth can be computed in some specified time or space bound by a given program or set of programs. This type of "awareness" could 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. [RSA,Me,GMR]). 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 remark that knowledge seems more appropriate than belief when trying to capture reasoning about cryptographic protocols.

The message that the reader should get from these examples is that the ability to place conditions on the awareness function provides a flexible tool for modelling various situations. Even greater flexibility can be attained once we incorporate time into the language; this is the subject of the next section.

5. Incorporating time

We can further extend this logic (and in fact all the other logics we have been discussing) by adding a relation, and a corresponding modal operator, to capture time. Formally, a model would now be a tuple $M = (S, \pi, \mathcal{A}_1, \ldots, \mathcal{A}_n, \mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{T})$, where $\mathcal{T}$ is a deterministic, serial relation; i.e. for all $s \in S$, there is a unique $t \in S$ such that $(s,t) \in \mathcal{T}$. Intuitively, $(s,t) \in \mathcal{T}$ if t describes the state of the world at the "next" time instant after s.[4] 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 $\mathcal{T}^*$ to be the reflexive, transitive closure of $\mathcal{T}$, that is, the binary relation on S defined by $(s,t) \in \mathcal{T}^*$ iff there exist states $s_0, \ldots, s_k$ such that $s = s_0$, $t = s_k$, and $(s_i, s_{i+1}) \in \mathcal{T}$ for $i < k$. More formally, we have:

$M,s \models \bigcirc\varphi$ if $M,t \models \varphi$ for (the unique) t such that $(s,t) \in \mathcal{T}$.

$M,s \models \Diamond\varphi$ if $M,t \models \varphi$ for some $t \in \mathcal{T}^*$.

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 relationship between belief (or knowledge), time, and awareness. When considering knowledge rather than belief, in some treatments (for example [Sa,Leh1]), 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

(*) $K_i(\bigcirc\varphi) \Rightarrow \bigcirc K_i\varphi$;

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. $\mathcal{T} \circ \mathcal{B}_i \subseteq \mathcal{B}_i \circ \mathcal{T}$.

It is easy to check that axiom (*) holds in all models

[4] 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. (HCJ).

that obey the restriction (**) (although it is still an open question whether (*) characterizes such models). As pointed out to us by Elias Thijsse, (*) 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, (*) is not often not a realistic assumption. People certainly forget! And (•*) seems to have rather unpleasant consequences for the decision procedure of the resulting logic (see Section 7).

Recall that one interpretation we gave the awareness function in the previous section 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:

(†) if $(s,t) \in \mathcal{T}$ then $\mathcal{A}_i(s) \subseteq \mathcal{A}_i(t)$ and

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

Intuitively, constraint (†) says that agent i's awareness never decreases over time, while (‡) says that i is eventually aware of every formula. In a model 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. As long as φ and $\varphi \Rightarrow \psi$ 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 φ and $\varphi \Rightarrow \psi$ 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 (‡) while retaining (†), 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.

6. A logic of local reasoning

Although the logic of general awareness discussed in the previous two sections 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, but without making 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 [Wi] notes that the two great theories physicists reason with are the theory of quantum phenomena and the theory of relativity. However [RB, p. 166], Wigner thinks that the two theories may 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 model for local reasoning* (or a *cluster model*) 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, with the restriction that if $s' \in T \in \mathcal{C}_i(s)$, then $T \in \mathcal{C}_i(s')$. It turns out that this restriction ensures that if an agent believes a fact, then he believes that he believes it (i.e., $B_i\varphi \Rightarrow B_iB_i\varphi$).[5] Intuitively, if $\mathcal{C}_i(s) = \{T_1, \ldots, T_k\}$, then in state s player i sometimes (depending perhaps on 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 model is equivalent to the models of the previous section, where we interpret $(s,t) \in \mathcal{B}_i$ exactly if $t \in T_s$.

An interesting special case of these models is one where in each frame of mind, an agent refuses to admit that he may occasionally be in another frame

S If we with to capture knowledge rather than belief, then we need to add the further rettrieUon that $ if a member of every member of $V_i(s)$.

of mind. This phenomenon can 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 $\{T\}$.[6] We call a cluster model satisfying this restriction a *narrow-minded cluster model*.

With this semantics we are not really trying to capture implicit and explicit belief. Rather, we are trying to capture *weak* and *strong* belief, where a weak belief is one that is true of some frame of mind (i.e. all the states in some cluster), while a strong belief is one that is true in all frames of mind. Accordingly, we now interpret $B_i\varphi$ as *player i weakly believes* φ, and use the new modal operator S_i to denote strong belief, so that $S_i\varphi$ means *player i strongly believes* φ. We define $\models$ for cluster models as follows:

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

It is easy to see from the semantic definitions given that weak belief is not closed under implication, but in this case the reason has nothing to do with awareness. $B_ip \wedge B_i(p \Rightarrow q) \wedge \sim B_iq$ 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.

More importantly for our purposes, note that an agent may now hold inconsistent (weak) beliefs: $B_ip \wedge B_i{\sim}p$ is satisfiable, since in one frame of mind 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. Of course, $S_ip \wedge S_i{\sim}p$ is also unsatisfiable.

In the narrow-minded cluster model, an agent will believe he is consistent (even if he is not) since 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. Indeed, 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: in the narrow-minded cluster model, $B_i(B_ip \wedge B_i(p \Rightarrow q) \Rightarrow B_iq)$ is also valid.

Note that in both the general and narrow-minded cluster mode] an agent's beliefs are closed under valid implication and agents believe all valid formulas. This is because we have assumed that agents can do perfect reasoning within each cluster. We can easily combine the ideas of the cluster model with those of the general awareness model to get a model where agents do not necessarily believe all valid formulas. The details are straightforward and left to the reader.

7. Decision procedures and complete axiomatizations

In the case of the classical logics of belief and knowledge, SS and KD45, it is known that the problem of deciding whether a formula is satisfiable is NP complete in the case of one player, and PSPACE complete if there is more than one player (see [HM] for a discussion of these results). Despite the apparent extra machinery we have introduced in our models, we can show that the decision procedures get no harder.

Theorem 7.1. *For Levesque's model of implicit and explicit belief, and the one-knower version of the logics of awareness, the logic of general awareness, and the narrow-minded version of the logic of local reasoning, the problem of deciding satisfiability of formulas is NP complete (and hence the problem of deciding validity is co-NP complete). For the many-knower versions of all these logics, the one-knower and many-knower versions of the logic of general awareness with time, and the unrestricted version of the logic of local reasoning the problem of deciding satisfiability and validity of formulas is PSPACE complete.*

We remark that once we add condition (•*) to the semantics of knowledge and time, things seem to get much worse. There the best-known results are a double-exponential decision procedure in the case of one knower; the problem for many knowers is still open.[7]

Using standard techniques of modal logic, we can also provide complete axiomatizations for all the logics we have discussed. We discuss a complete axiomatization for the logic of general awareness here to show how the usual axioms of belief must be modified. Axiomatizations for the other logics we have discussed and further details of proofs will appear in the full paper.

Recall that in this logic we have $B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$, and L_i acts like the classical belief operator. It is well-known (cf. [Ch,FV,HM]) that belief can be

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

[7] The proof in the case of one knower is a modification of the techniques of [Leh1] In that paper, Lehmann also claims a double-exponential decision procedure for the logic of knowledge and time with many knowers, but his proof techniques seem to fail [Leh2].

axiomatized as follows:

1. All substitution instances of propositional tautologies.
2. $L_i\varphi \Rightarrow L_iL_i\varphi$ ("Introspection of positive belief").
3. $\sim L_i\varphi \Rightarrow L_i{\sim}L_i\varphi$ ("Introspection of negative belief").
4. $L_i\varphi_1 \wedge L_i(\varphi_1 \Rightarrow \varphi_2) \Rightarrow L_i\varphi_2$ ("What player i believes is closed under modus ponens").
5. $\sim L_i(\mathit{false})$ ("Player i does not believe a contradiction").

There are two inference rules: (1) from φ_1 and $\varphi_1 \Rightarrow \varphi_2$ infer φ_2 (modus ponens) and (2) from φ infer $L_i\varphi$ ("belief of tautologies").[8]

If we simply add to this collection of axioms two more axioms, namely $A_i(\mathit{false})$ ("player i is aware of the formula *false*") and $B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$ ("explicit belief is equivalent to implicit belief plus awareness"), then we get a complete axiomatization of the logic of general awareness.

Theorem 7.2. *The axiom system described above is a sound and complete axiomatization of the logic of general awareness.*

Even more insight into the logic is obtained if we consider the fragment of the logic where we can only discuss explicit belief, and not implicit belief (i.e, if we consider the language without the L_i operators). Consider the following axioms schemas and rules of inference.

1. All substitution instances of propositional tautologies.
2. $B_i\varphi \wedge A_iB_i\varphi \Rightarrow B_iB_i\varphi$
3. $\sim B_i\varphi \wedge A_i(\sim B_i\varphi) \Rightarrow B_i{\sim}B_i\varphi$
4. $B_i\varphi_1 \wedge B_i(\varphi_1 \Rightarrow \varphi_2) \wedge A_i\varphi_2 \Rightarrow B_i\varphi_2$
5. $\sim B_i(\mathit{false})$
6. $A_i(\mathit{false})$.
7. $B_i\varphi \Rightarrow A_i\varphi$

There are three inference rules: (1) from φ and $\varphi \Rightarrow \psi$ infer ψ, (2) from ψ infer $A_i\psi \Rightarrow B_i\psi$, and (3) from $B_i\varphi_1 \wedge \ldots \wedge B_i\varphi_n \wedge \varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow \psi$ infer $B_i\varphi_1 \wedge \ldots B_i\varphi_n \wedge A_i\psi \Rightarrow B_i\psi$. (We can view (2) as a special case of (3) with $n = 0$.) Both (2) and (3) can be viewed as rules of "limited beliefs of tautologies".

Note how the crucial differences between implicit and explicit belief are reflected in the axiom system, particularly axioms 2, 3, and 4 and the rules of limited belief of tautologies. In all these cases an agent must be aware of the relevant formula before he believes it. Of course, similar remarks hold if we replace belief by knowledge. Note that axiom 2 indicates how, according to de Chardin, an animal may know, but not know that it knows, while axiom 3 indicates how an agent may be "so dumb that he doesn't even know that he doesn't know p".

Theorem 7.3. *The axiom system above is a sound and complete axiomatization for the language of general awareness without the implicit belief operator.*

It is also straightforward to axiomatize some of the restrictions on awareness mentioned in Section 4. For example, if the order of presentation of the conjuncts does not matter, we have $A_i(\varphi \wedge \psi) \equiv A_i(\psi \wedge \varphi)$; if an agent is always simultaneously aware of a formula and its negation, we have $A_i\varphi \equiv A_i{\sim}\varphi$. If we take awareness to be closed under subformulas, then this can be captured axiomatically by adding the axiom schemas $A_i(\sim\varphi) \Rightarrow A_i\varphi$, $A_i(\varphi \wedge \psi) \Rightarrow A_i\varphi \wedge A_i\psi$, $A_i(B_j\varphi) \Rightarrow A_i\varphi$, $A_i(L_j\varphi) \Rightarrow A_i\varphi$, and $A_i(A_j\varphi) \Rightarrow A_i\varphi$. With these additional axioms, it can be shown that an agent's beliefs are closed under implication (that is, whenever both φ and $\varphi \Rightarrow \psi$ are believed, then ψ is believed), although agents still do not believe all valid formulas. By changing $\Rightarrow$ to $\equiv$ in these axioms, we can capture a notion of awareness generated by a set of primitive propositions. We note that in the case of awareness generated by a set of primitive propositions, a number of the axioms and rules simplify. In particular, we can omit the clauses involving awareness in axioms 2, 3, and 4, and omit rule of inference (3) altogether. Further details can be found in the full paper.

8. 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 are currently investigating 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. We hope to report on these issues in a future paper.

[8] Note that in the case of knowledge, rather than belief, we replace L_i by K_i, and replace the axiom $\sim L_i(\mathit{false})$ by $K_i\varphi \Rightarrow \varphi$ ("Whatever player i knows is true").

9. Acknowledgements

We would like to thank Elias Thijsse for his detailed reading of the paper and numerous pertinent comments and criticisms. We would like to thank Kurt Konolige, Hector Levesque, Yoram Moses, Peter van Emde Boas, Moshe Vardi, and Ed Wimmers, for their helpful remarks.

10. References

[AB] A. R. Anderson and N. D. Belnap, *Entailment, the Logic of Relevance and Necessity,* Princeton University Press (1975).

[BI] A. Borgida and T. Imielinski, Decision making in committees - a framework for dealing with inconsistency and non-monotonicity, *Proc. Nonmonotonic Reasoning Workshop,* (1984).

[Ch] B. F. Chellas, *Modal Logic,* Cambridge University Press (1980).

[Cr] M. J. Cresswell, *Logics and Languages,* Methuen and Co. (1973).

[Do] J. Doyle, A society of mind, *Proc. International Joint Conference on Artificial Intelligence (IJCAI),* 1983.

[Eb] R. A. Eberle, A logic of believing, knowing and inferring, *Synthese* 26 (1974), pp. 356-382.

[FHV] R. Fagin, J. Y. Halpern, and M. Y. Vardi, A model-theoretic analysis of knowledge, *Proc. 25th IEEE Symposium on Foundations of Computer Science, West Palm Beach, Florida* (1984), pp. 268-278.

[FV] R. Fagin and M. Y. Vardi, An internal semantics for modal logic, *Proc. of the 17th Symposium on Theory of Computing* (1985), pp. 305-315.

[GMR] S. Goldwasser, S. Micali, and C. Rackoff, The knowledge complexity of interactive proof-systems, *Proc. of the 17th Symposium on Theory of Computing* (1985), pp. 291-304.

[HM] J. Y. Halpern and Y. O. Moses, A guide to the modal logics of knowledge and belief, to appear as an IBM Research Report (1985).

[Hi] J. Hintikka, Impossible possible worlds vindicated, *J. Philosophical Logic* 4 (1975), pp. 475-484.

[HC] G.E. Hughes and M.J. Cresswell, *An Introduction to Modal Logic,* Methuen, London (1968).

[Ko] K. Konolige, Belief and incompleteness, SRI Artificial Intelligence Note 319, SRI International, Menlo Park (1984). pp. 377-381.

[Kr] S. Kripke, Semantical analysis of modal logic, *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik* 9 (1963), pp. 67-96.

[Lehi] D.J. Lehmann, Knowledge, common knowledge, and related puzzles, in "Proceedings of the Third Annual ACM Conference on Principles of Distributed Computing", 1984, pp. 62-67.

[Leh2] D.J. Lehmann, private correspondence.

[Lev1] H. J. Levesque, A logic of implicit and explicit belief, *Proc. Nall Conf. on Artificial Intelligence* (1984), pp. 198-202; a revised and expanded version appears as FLAIR Technical Report #32, (1984).

[Lev2] H. J. Levesque, Global and local consistency and completeness of beliefs, in preparation.

[Lu] J. Lukasiewicz, O logice trojwartosciowej (On three-valued logic), *Ruch Filozoficzny* 5 (1920), pp. 169-171.

[Me] M. J. Merritt, Cryptographic protocols, Ph.D. Thesis, Georgia Institute of Technology, 1983.

[Mi] M. Minsky, Plain talk about neurodevelopmental epistemology, *Proc. 5th Int. Joint Conf. on AI* (1977), pp. 1083-1092.

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

[Ra] V. Rantala, Impossible worlds semantics and logical omniscience, *Acta Philosophica Fennica* 35 (1982), pp. 106-115.

[RB] N. Rescher and R. Brandom, *The Logic of Inconsistency,* Rowman and Littlefield (1979).

[RSA] R. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Comm. of the ACM,* 21:2 (1978), pp. 120-126.

[Sa] M. Sato, A study of Kripke-style methods of some modal logics by Gentzen's sequential method, *Publications of the Research Institute for Mathematical Sciences, Kyoto University,* 13:2, 1977.

[Se] K. Segerberg, *An essay on classical modal logic,* Uppsala, Philosophical Studies (1972).

[St] R. Stalnaker, *Inquiry,* M.I.T. Press, 1985.

[Va] M. Y. Vardi, On epistemic logic and logical omniscience, unpublished manuscript.

[Wi] Eugene P. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, *Comm. on Pure and Applied Math.* 13 (1960), pp. 1-14.

[Za] W. Zadrozny, Explicit and implicit beliefs, a solution of a problem of H. Levesque, unpublished manuscript, 1985.