A Logic of Implicit and Explicit Belief

Hector J. Levesque

Fairchild Laboratory for Artificial Intelligence Research
4001 Miranda Avenue
Palo Alto, California 94304

ABSTRACT

As part of an on-going project to understand the foundations of Knowledge Representation, we are attempting to characterize a kind of belief that forms a more appropriate basis for Knowledge Representation systems than that captured by the usual possible-world formalizations begun by Hintikka. In this paper, we point out deficiencies in current semantic treatments of knowledge and belief (including recent syntactic approaches) and suggest a new analysis in the form of a logic that avoids these shortcomings and is also more viable computationally.

The kind of belief that underlies terms in AI such as "Knowledge Representation" or "knowledge base" has never been adequately characterized.[1] As we discuss below, the major existing formal model of belief (originated by Hintikka in [1]) requires the beliefs of an agent to be closed under logical consequence, and thus can place unrealistic computational demands on his reasoning abilitites. Here we describe and formalize a weaker sense of belief that is much more attractive computationally and forms a more plausible foundation for the service to be provided by a Knowledge Representation utility. This formalization is done in the context of a logic of belief that has a truth-based semantic theory (like the possible-world approach but unlike its recent syntactic competitors). This logic is also shown to have connections to relevance logic and, in a certain sense, to subsume it.

1. Logical Omniscience & Possible Worlds

A recurring problem in the modelling of belief or knowledge is what has been called in [2] *logical omniscience*. In a nutshell, all formalizations of belief based on a possible-world semantics suffer from the fact that at any given point, the set of sentences considered to be believed is closed under logical consequence. It is simply built into the logic that if α is believed and α logically implies β, then β is believed as well. Apart from the fact that this does not allow for a resource-limited agent who might fail to draw any connection between α and β, this has at least three other serious drawbacks from a modelling point of view:

1. Every valid sentence must be believed.
2. If two sentences are logically equivalent, then one must be believed if the other is.
3. If a sentence and its negation are both believed, then so must *every* sentence.

Any one of these might cause one to reject a possible-world formalization as unintuitive at best and completely unrealistic at worst.

There is, however, a much more reasonable way of interpreting the possible-world characterization of belief. As discussed in [3], instead of taking logical omniscience as an idealization (or heuristic) in the modelling of the beliefs of an agent, we can understand it to be dealing realistically with a different though related concept, namely, what is *implicit* in what an agent believes. For example, if an agent imagines the world to be one where α is true and if α logically implies β, then (whether or not he realizes it) he imagines the world to be one where β also happens to be true. In other words, if the world the agent believes in satisfies α, then it must also satisfy β. Under this interpretation, we examine not what an agent believes directly, but *what the world would be like if what he believed were true.* There are often very good reasons for examining the consequences of what an agent believes even if the agent himself has not yet appreciated those consequences.

If the proper understanding of a possible-world semantics is that it deals not with what is believed, but what is true given what is believed, what then is an appropriate semantics for dealing with the actual beliefs of an agent? Obviously, we need a concept other than the one formalized by possible worlds. If we use the terminology that a sentence is *explicitly* believed when it is actively held to be true by an agent and *implicitly* believed when it follows from what is believed, then what we want is a formal logical language that includes two operators, B and L: $B\alpha$ will be true when α is explicitly believed while $L\alpha$ will be true when α is implicit in what is believed. While a possible-world semantics (like that of [1] or [4]) is appropriate for dealing with the latter concept, the goal of this paper is to present one for the former.

2. The Syntactic Approach

When talking about what an agent actually believes, we want to be able to distinguish between believing only α and $(\alpha \supset \beta)$ on the one hand, and believing α, $(\alpha \supset \beta)$ and β, on the other. While the picture of the world is the same in both cases, only the second involves realizing that β is true. This is somewhat of a problem semantically, since the two sets of beliefs are true in precisely the same possible worlds and so, in some sense, seman-

[1] Because what is represented in a knowledge base is typically not required to be *true*, to be consistent with most philosophers and computer scientists, we are calling the attitude involved here "belief" rather than "knowledge".

tically indistinguishable. This might suggest that any realistic semantics for belief will have to include (something isomorphic to) a set of sentences to distinguish between the two belief sets above. The usual way to interpret a sentence like $L\alpha$ in a standard Kripke framework is to have a model structure that contains a set of possible worlds, an accessibility relation and other things. It appears that to interpret a sentence like $B\alpha$, a model structure will have to contain an explicit set of sentences. This is indeed what happens in the formalizations of belief of [5] and [6] that share our goal of avoiding logical omniscience. A slightly more sophisticated approach is that of [7] where the semantic structure contains only an initial set of sentences (representing a base set of beliefs) and a set of logically sound deductive rules for obtaining new derived beliefs. Logical omniscience is avoided there by allowing the deductive rules to be logically incomplete. With or without deductive rules, I will refer to this approach to modelling belief as the *syntactic* approach since syntactic entities have to be included within the semantic structures.

Apart from this perhaps ill-advised mixture of syntax and semantics, the syntactic approach suffers from a serious defect that is the opposite of the problem with possible worlds. A possible-world semantics is, in some sense, too *coarse-grained* to model belief in that it cannot distinguish belief sets that logically imply the same set of sentences. The syntactic approach, on the other hand, is too *fine-grained* in that it considers any two sets of sentences as distinct semantic entities and, consequently, different belief sets.

To see why this a problem, consider, for example, the disjunction of α and β. There is no reason to suppose that

$$B(\alpha \vee \beta) \equiv B(\beta \vee \alpha)$$

would be *valid* given a syntactic understanding of B since $(\alpha \vee \beta)$ may be in the belief set while $(\beta \vee \alpha)$ may not.[2] The trouble with this is that if we consider intuitively what

"It is believed that either α or β is true."

is saying, the order seems to be completely irrelevant. It is almost an accident of lexical notation that we had to choose one of the disjuncts to go first. Yet, the syntactic approach makes the left to right order of disjuncts *semantically* significant in that we can believe one ordering but fail to believe the other.

The obvious counter to this is that the logic of the syntactic approach has to be embellished to avoid these spurious syntactic distinctions. For example, we might insist as part of the semantics that to be well-formed, any belief set containing $(\alpha \vee \beta)$ must also contain $(\beta \vee \alpha)$ (or, for Konolige, the obvious deduction rule must be present). The trouble with this kind of constraint is that it is semantically unmotivated. For example, should we also insist that any set containing $\neg\neg\alpha$ must also contain α? Should every belief set containing α and β also contain $(\alpha \wedge \beta)$? Should every belief set contain the "obvious" tautologies such as $(\alpha \supset \alpha)$? Where do we stop? Clearly, it would be preferable to have a semantics where restrictions such as these *follow* from the meaning of $B\alpha$ and not the other way around. In other words, we want a semantics (like that of possible worlds) that is based on some concept of *truth* rather than on a collection of *ad hoc* restrictions to sets of sentences. Ideally, moreover, the granularity of the semantics should lie somewhere between the syntactic and the possible-world approaches so that different sets of sentences can represent the same beliefs without requiring that all logically equivalent sets do so. We now show that there is a reasonably intuitive semantics for belief that has these properties.

[2] In Konolige's system, one disjunction may be deducible while the other may not.

3. Situations

On closer examination, the reason the possible-world approach to belief or knowledge leads to logical omniscience is that beliefs are characterized completely by a set of possible worlds (namely, those that are accessible from a given possible world). Intuitively, these possible worlds are to be thought of as the full range of what the agent thinks the world might be like. If he only believes that p is true, the set of worlds will be all those where p is true: some, for example, where q is true, others, where q is false. However, because sentences which are tautologies will also be true in all these possible worlds, the agent is thought of as believing them just as if they were among his active beliefs. In terms of the possible worlds, there is no way to distinguish p from these tautologies.

One way to avoid all these tautologies is to to make this notion of what an agent thinks the world is like be more relevant to what he actually believes. This can be done by replacing the possible worlds by a different kind of semantic entity that does not necessarily deal with the truth of all sentences. In particular, sentences not relevant to what an agent actually believes (including some tautologies) need not get a truth value at all. Following [8] (but not too closely), we will call this sort of partial possible world a *situation*. Roughly speaking, a situation may support the truth of some sentences and the falsity of others, but may fail to deal with other sentences at all.

For example, consider the situation of me sitting at my terminal at work. We might say that this situation supports the fact that I'm at work, that somebody is at my terminal, that there is either a terminal or a book at my desk, and so on. On the other hand, it does not support the contention that my wife is at home, that she is not out shopping, or even that she is at home or not at home. Although the latter is certainly *true*, me sitting at my terminal does not deal with it one way or another.

One way of thinking about situations is as generalizations of possible worlds where not every sentence in a language is required to have a truth value. Conversely, we can think of possible worlds as those limiting cases of situations where every sentence does have a truth value. Indeed, the concept of a possible world being *compatible* with a situation is intuitively clear: every sentence whose truth is supported by the situation should come out true in that possible world and every sentence whose falsity is supported should come out false. Again drawing from [8], we will also allow for *incoherent* situations with which no possible world is compatible. These are situations that (at least seem to) support both the truth and falsity of some sentence. From the point of view of modelling belief, these are very useful since they will allow an agent to have an incoherent picture of the world.

The "trick", then, that underlies the logic of belief to follow is to identify explicit belief with a *set of situations* rather than possible worlds. Before examining the formal details, there is one point to make. Traditional logics of knowledge and belief

have dealt not only with world knowledge but also with meta-knowledge, that is, knowledge about knowledge. To be able to deal with this in our case is somewhat of a problem since we would have to deal with a whole raft of questions about what is believed about what is explicitly or implicitly believed. For example, even without assuming that everything believed is true, it is not clear whether or not $B(L\alpha \supset \alpha)$ should be valid. For reasons given in [3], $L(L\alpha \supset \alpha)$ ***should*** be valid even if belief does not, in general, imply truth. Instead of trying to settle all of these questions here and now, we will ignore them completely. The language below will simply not contain any sentences where a B or a L appears within the scope of another. This will simplify the semantics immensely while still illustrating how the two concepts can co-exist naturally.

4. A Formal Semantics

The language we are considering (call it **L**) is formed in the obvious way from a set of atomic sentences **P** using the standard connectives $\vee$, $\wedge$, and $\neg$ for disjunction, conjunction, and negation respectively, and two unary connectives B and L. Only regular propositional sentences (without a B or a L) can occur within the scope of these last two connectives. We assume that other connectives such as $\supset$ and $\equiv$ can be understood in terms of the original ones.[3]

Sentences of **L** are interpreted semantically in terms of a *model structure* $\langle S, \mathcal{B}, \mathcal{T}, \mathcal{F} \rangle$ where S is a set, $\mathcal{B}$ is a subset of S, and both $\mathcal{T}$ and $\mathcal{F}$ are functions from **P** (the atomic sentences) to subsets of S. Intuitively, S is the set of all situations with $\mathcal{B}$ being those situations that could be the actual one according to what is believed. For any atomic sentence p, $\mathcal{T}(p)$ are the situations that support the truth of p and $\mathcal{F}(p)$ are those that support the falsity of p.

To deal with the possible worlds compatible with a situation in a model structure, we define $\mathcal{W}$ by the following:

$\mathcal{W}(s) = \{s' \in S \mid$ for every $p \in \mathbf{P}$,
a) s' is a member of exactly one of $\mathcal{T}(p)$ and $\mathcal{F}(p)$,
b) if s is a member of $\mathcal{T}(p)$, then so is s',
and *c)* if s is a member of $\mathcal{F}(p)$, then so is s'.$\}$

The first condition aboves guarantees that s' will be a possible world, while the last two guarantee compatibility. Also, for any subset S^* of S, we will let $\mathcal{W}(S^*)$ mean the union of all $\mathcal{W}(s)$ for every s in S^*.

Given a semantic structure $\langle S, \mathcal{B}, \mathcal{T}, \mathcal{F} \rangle$, we can define the *support relations* $\models_T$ and $\models_F$ holding between situations and sentences of **L**. Intuitively, $s \models_T \alpha$ when s supports the truth of α, and $s \models_F \alpha$ when s supports the falsity of α. More formally, we have the following:

$\models_T$ and $\models_F \subseteq S \times \mathbf{L}$ and are defined by

1. $s \models_T p$ iff $s \in \mathcal{T}(p)$.
   $s \models_F p$ iff $s \in \mathcal{F}(p)$.
2. $s \models_T (\alpha \vee \beta)$ iff $s \models_T \alpha$ or $s \models_T \beta$.
   $s \models_F (\alpha \vee \beta)$ iff $s \models_F \alpha$ and $s \models_F \beta$.
3. $s \models_T (\alpha \wedge \beta)$ iff $s \models_T \alpha$ and $s \models_T \beta$.
   $s \models_F (\alpha \wedge \beta)$ iff $s \models_F \alpha$ or $s \models_F \beta$.
4. $s \models_T \neg\alpha$ iff $s \models_F \alpha$.
   $s \models_F \neg\alpha$ iff $s \models_T \alpha$.
5. $s \models_T B\alpha$ iff for every s' in $\mathcal{B}$, $s' \models_T \alpha$.
   $s \models_F B\alpha$ iff $s \not\models_T B\alpha$.
6. $s \models_T L\alpha$ iff for every s' in $\mathcal{W}(\mathcal{B})$, $s' \models_T \alpha$.
   $s \models_F L\alpha$ iff $s \not\models_T L\alpha$.

If s is an element of $\mathcal{W}(S)$ (i.e. s is a possible world), then if $s \models_T \alpha$, we say that α ***is true at*** s and otherwise that α ***is false at*** s. Thus, as to be expected, a sentence is true iff it is not false iff its negation is false. Finally, we say that α ***is valid*** and write $\models \alpha$ provided that for any model structure $\langle S, \mathcal{B}, \mathcal{T}, \mathcal{F} \rangle$ and any s in $\mathcal{W}(S)$, α is true at s. The satisfiablitity of a sentence (or of a set of sentences) can be defined analogously. This completes the semantics of **L**.

While space precludes a lengthy examination of the properties of **L**, here are the major highlights. First of all, **L** handles its standard propositional subset correctly in that all instances of propositional tautologies are valid and, moreover, any sentence not containing a B or L is valid iff it is a standard tautology.

As for implicit belief, it is easy to see that all tautologies are implicitly believed and that it is closed under implication. In other words, we have

If $\models \alpha$ (where α is propositional), then $\models L\alpha$ and
$\models (L\alpha \wedge L(\alpha \supset \beta)) \supset L\beta$.

Equally important, the sentence $(B\alpha \supset L\alpha)$ is valid, meaning that everything that is explicitly believed is an implicit belief. In fact, if a sentence is a logical consequence[4] of what is believed, then it is implicitly believed. Unfortunately, the converse does not hold since in some interpretations, there may be sentences that are true in the right set of possible worlds without being implied by what is believed. For example, if a sentence is ***necessarily true*** then it will be an implicit belief—even if it is not logically valid—a generic problem with the possible-world semantics for knowledge and belief that seems to have gone unnoticed in the literature. We should not be *too* concerned about this, however, since it does not affect either the valid or the satisfiable sentences of **L**, but only whether or not certain ***infinite*** sets of sentences are satisfiable.[5]

Of course, the major issue here is how the B operator behaves. Before examining the valid sentences containing B, it is worth considering some satisfiable sets of sentences that show that belief does not suffer from logical omniscience. The following sets are all satisfiable:

1. $\{Bp, B(p \supset q), \neg Bq\}$ This shows that beliefs are not closed under implication.

[3] We may eventually want a special implication operator, especially for sentences that are objects of belief.

[4] A sentence α is a logical consequence of a set L^* of sentences iff $L^* \cup \{\neg\alpha\}$ is unsatisfiable.

[5] There is, moreover, a fairly simple way to eliminate the problem of non-logical necessary truths always being implicitly believed. Call a model structure *expansive* if for any set of atomic sentences, there is a possible world in the structure such that the atomic sentences it supports is precisely that set. Now while there are certainly model structures that are not expansive, it can be shown that the validity or satisfiability of a sentence would not change if these were defined in terms of expansive structures only. With this definition, moreover, a sentence would indeed be implicitly believed if *and only if* it was logically implied by what was believed.

2. $\{\neg B(p \vee \neg p)\}$ A valid sentence need not be believed.

3. $\{Bp, \neg B(p \wedge (q \vee \neg q))\}$ A logical equivalent to a belief need not be believed.

4. $\{Bp, B\neg p, \neg Bq\}$ Beliefs can be inconsistent without every sentence being believed.

The above sets show what freedom the logic allows in terms of belief; to demonstrate that the logic does impose reasonable constraints on belief, we must look at the valid sentences of **L**. We will present these in terms of a proof theory for **L** that is both sound and complete with respect to the above semantics. The important point, however, is that unlike the syntactic approach, these constraints follow from the semantics. The only reason to consider a proof theory here is that it does provide an elegant and vivid way to examine the valid sentences of **L** (especially those using B).[6]

5. A Proof Theory

The proof theory of **L** must begin with a propositional basis of some sort to guarantee that all tautologies are present. The simplest way to do this is to have a single rule of inference, *Modus Ponens*, and the usual three axioms that can be found in any elementary logic textbook. To this basic system we will adjoin a collection of new axioms for implicit and explicit belief but no new rules of inference.

The appropriate axioms for implicit belief should make sure that it contains all tautologies and all beliefs and is closed under implication. This can be achieved with three axiom schemata:

1. $L\alpha$, where α is a tautology.

2. $(B\alpha \supset L\alpha)$.

3. $L\alpha \wedge L(\alpha \supset \beta) \supset L\beta$.

For explicit belief, on the other hand, we have to dream up a set of axioms stating what has to be believed when something else is. In other words, we need a set of axioms of the form $(B\alpha \supset B\beta)$, for various α and β. Remarkably enough, this work has already been done for us in what is called *relevance logic* [9]. This logic deals with a relationship between pairs of sentences called *entailment* that is a proper subset of logical implication. Entailment is based on the intuition that the antecedent of an implication should be relevant to the consequent. As it turns out, entailment and belief are very closely related, as the following key result attains:

Theorem 1: $\models (B\alpha \supset B\beta)$ *iff* α *entails* β.

The proof of this theorem[7] is based on a correspondence between our semantics of situations and a semantics of four truth-values described in [11]. What this tells us is that **L** contains relevance logic as a subpart: questions of entailment can be reduced to questions of belief in **L**. Moreover, we get this relevance logic without having to give up classical logic and the normal interpretation of $\supset$ and the other connectives.

[6] We could imagine constructing a decision procedure for **L** directly from the above without even passing through a proof theory at all. Such a decision procedure, after all, is what counts when building a system that reasons with **L**.

[7] Proofs of this and the two other quoted theorems can be found in [10], a slightly revised version of this paper.

So all that is needed to characterize the constraints satisfied by belief is to apply a set of axioms for entailment in relevance logic to belief. One such set given in [9] is the following:

4. $B(\alpha \wedge \beta) \equiv B(\beta \wedge \alpha)$.
   $B(\alpha \vee \beta) \equiv B(\beta \vee \alpha)$.

5. $B(\alpha \wedge (\beta \wedge \gamma)) \equiv B((\alpha \wedge \beta) \wedge \gamma)$.
   $B(\alpha \vee (\beta \vee \gamma)) \equiv B((\alpha \vee \beta) \vee \gamma)$.

6. $B(\alpha \wedge (\beta \vee \gamma)) \equiv B((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$.
   $B(\alpha \vee (\beta \wedge \gamma)) \equiv B((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$.

7. $B\neg(\alpha \vee \beta) \equiv B(\neg\alpha \wedge \neg\beta)$.
   $B\neg(\alpha \wedge \beta) \equiv B(\neg\alpha \vee \neg\beta)$.

8. $B\neg\neg\alpha \equiv B\alpha$.

9. $B\alpha \wedge B\beta \equiv B(\alpha \wedge \beta)$.
   $B\alpha \vee B\beta \supset B(\alpha \vee \beta)$.

This particular axiomatization states that belief must respect properties of the logical operators such as commutativity, associativity, distibutivity, De Morgan's laws and double negation. Nothing in these axioms forces *all* the logical consequences of what is believed to be believed (as in axioms 1 and 3, above, for implicit belief), although each one forces *some* consequences to be believed (e.g., by axiom 8, a double negation of a sentence must be believed if the sentence itself is).

Another way to understand these axioms (except for the very last one) is as constraints on the individuation of beliefs. For example, $(\alpha \vee \beta)$ is believed iff $(\beta \vee \alpha)$ is because these are two lexical notations for the *same* belief. In this sense, it is not that there is an automatic inference from one belief to another, but rather two ways of describing a single belief.

This, in itself, does not *justify* the axioms, however. It is easy to imagine logics of belief that are different from this one, omitting certain of the above constraints or perhaps adding additional ones. Indeed, there is not much to designing a proof theory with any collection of constraints on belief. The interesting fact about *this* particular set of axioms, however, is that it corresponds so nicely to an independently motivated semantic theory. Specifically, we have the following result:

Theorem 2: (Soundness and Completeness)
A sentence of **L** *is a theorem of the above logic iff it is valid.*

Furthermore, and perhaps most importantly, the logic of **L** has very attractive computational properties as well, which we now turn to.

6. The Payoff

What does this new logic of belief buy us? One thing is a language that can be used to formally reason about the beliefs of other agents without assuming logical omniscience. If we imagine a system planning speech acts as in [12], we can represent what it knows about the beliefs of another as a theory in **L**. It could then plan to remind someone of something he only believes implicitly. Similarly, it could take someone through certain steps of an argument or proof, at each stage pointing out implications of the other agent's beliefs.

There are any number of ways to mechanize the necessary reasoning in **L**. One currently fashionable method involves translating everything into first-order logic and running a resolution

theorem-prover over the results. This would involve the usual encoding of sentences of **L** as terms and characterizing either its validity or provability (or both) using a first-order theory. Just doing this, however, would miss a very important feature of **L**, namely that calculating propositional beliefs is much easier than doing general propositional reasoning.

Consider, in particular, the role of a logical Knowledge Representation system (such as KRYPTON [13]) that is given as a knowledge base (or KB) a finite set of sentences in some language. What a knowledge-based system using this KB (such as a robot) will be interested in is whether or not some proposition is true of the application domain (e.g. "Is it raining outside?"). The ideal way of answering this kind of questions is *yes* if the question follows from what is in the KB, *no* if its negation does and *unknown* otherwise. The sad fact of the matter, however, is that for all but extremely simple languages (including some without quantifiers) this question-answering is computationally intractable. This might be tolerable if the kind of question you ask is an open problem in mathematics where you are willing to stop and redirect the theorem-prover with problem-specific heuristics if it seems to be thrashing. If, on the other hand, a robot is trying to decide whether or not to use an umbrella, and calls a Knowledge Representation system utility as a subroutine, this kind of behaviour is unacceptable.

A possible solution to the problem is for the Knowledge Representation system to manage what is explicitly *believed* rather than its implications. In those cases where a question cannot be answered directly on the basis of what is believed, the robot can decide to try to figure out the answer by determining the implications of what it believes. Moreover, new facts can be sought and the question can even be abandoned it it becomes too expensive to pursue (e.g. the robot can decide to bring its umbrella just to be safe). The point is that this more general form of reasoning can be controlled very carefully depending on the situation since it is no longer just a subroutine call to a Knowledge Representation system. The robot can, in fact, *plan* to figure something out just as it would plan any other activity.

This is all very speculative, of course. How do we know, for example, that it is any easier to calculate what is believed rather than its implications? There is, fortunately, fairly strong evidence for this, at least in the propositional case:

Theorem 3: *Suppose KB and α are propositional sentences in conjunctive normal form. Determining if KB logically implies α is co-NP-complete but determining if KB entails α has an $O(mn)$ algorithm, where $m = |KB|$ and $n = |\alpha|$.*

Corollary 4: *Assume KB and α are as above. Then, in the worst case, deciding if*

a) $\models (B_{KB} \supset L\alpha)$ *is very difficult.*

b) $\models (B_{KB} \supset B\alpha)$ *is relatively easy.*

What this amounts to is that if we consider answering questions of a given fixed size, the time it takes to calculate what the KB believes will grow *linearly* at worst with the size of the KB, but the time it takes to calculate the implications of what the KB believes will grow *exponentially*[5] at worst with the size of the KB.

[5]More precisely, it will grow faster than any polynomial function, unless **P** equals **NP**.

Returning now to the formal modelling of the beliefs of other agents, the reason we would not want to simply run an untuned resolution theorem-prover over encodings of sentences of **L** is that we would lose the opportunity to exploit the computational tractability of belief.

Again, it is not so much that our logic is the *only* one to capture a semantically and computationally respectable notion of belief. What it demonstrates, however, is first, that it is possible to move away from closure under classical implication without espousing the syntactic approach and giving up semantics altogether, and second, that there is hope for a non-trivial domain-independent Knowledge Representation deductive service. Of course, it remains to be seen whether these advantages can be preserved for a language that includes meta-knowledge and quantifiers. Discovering appropriate semantics and decision procedures in these cases remains a difficult open problem.

ACKNOWLEDGEMENTS

This work was done as part of the KRYPTON project at Fairchild and I am indebted to its other members, Ron Brachman, Richard Fikes, Peter Patel-Schneider, and Victoria Pigman, as well as to David Israel of BBN, Joe Halpern and the other participants of the Knowledge Seminar at IBM San Jose, and to the Best Western family of hotels.

REFERENCES

[1] Hintikka, J., *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, 1962.

[2] Hintikka, J., Impossible Possible Worlds Vindicated, *Journal of Philosophical Logic*, **4**, 1975, 475–484.

[3] Levesque, H. J., Foundations of a Functional Approach to Knowledge Representation, *Artificial Intelligence*, forthcoming.

[4] Moore, R. C., Reasoning about Knowledge and Action, Technical Note 181, SRI International, Menlo Park, 1980.

[5] Moore, R. C. and Hendrix, G., Computational Models of Beliefs and the Semantics of Belief-Sentences, Technical Note 187, SRI International, Menlo Park, 1979.

[6] Eberle, R. A., A Logic of Believing, Knowing and Inferring, *Synthese* **26**, 1974, 356–382.

[7] Konolige, K., A Deduction Model of Belief, Ph. D. Thesis, Computer Science Department, Stanford University, in preparation.

[8] Barwise, J. and Perry, J., *Situations and Attitudes*, Bradford Books, Cambridge, MA, 1983.

[9] Anderson, A. R. and Belnap, N. D., *Entailment, The Logic of Relevance and Necessity*, Princeton University Press, 1975.

[10] Levesque, H. J., A Logic of Implicit and Explicit Belief, Fairchild Laboratory for Artificial Intelligence Research, Technical Report, in preparation.

[11] Belnap, N. D., A Useful Four-Valued Logic, in G. Epstein and J. M. Dunn (eds.), *Modern Uses of Multiple-Valued Logic*, Reidel, 1977.

[12] Perrault, C. R. and Cohen, P. R., Elements of a Plan-Based Theory of Speech Acts, *Cognitive Science* **3**, 1979, 177–212.

[13] Brachman, R. J., Fikes, R. E., and Levesque, H. J., KRYPTON: A Functional Approach to Knowledge Representation, *IEEE Computer*, **16** (10), 1983, 67–73.