An Operational Semantics for Knowledge Bases*

Ronald Fagin
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120–6099
fagin@almaden.ibm.com

Joseph Y. Halpern†
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120–6099
halpern@almaden.ibm.com

Yoram Moses‡
Deparment of Applied Math. and CS
The Weizmann Institute of Science
76100 Rehovot, Israel
yoram@wisdom.weizmann.ac.il

Moshe Y. Vardi§
Department of Computer Science
Rice University
Houston, TX 77251-1892
vardi@cs.rice.edu

Abstract

The standard approach in AI to knowledge representation is to represent an agent's knowledge symbolically as a collection of formulas, which we can view as a knowledge base. An agent is then said to know a fact if it is provable from the formulas in his knowledge base. Halpern and Vardi advocated a model-theoretic approach to knowledge representation. In this approach, the key step is representing the agent's knowledge using an appropriate semantic model. Here, we model knowledge bases operationally as multi-agent systems. Our results show that this approach offers significant advantages.

Introduction

The standard approach in AI to knowledge representation, going back to McCarthy **?**, is to represent an agent's knowledge symbolically as a collection of formulas, which we can view as a *knowledge base*. An agent is then said to know a fact if it is *provable* from the formulas in his knowledge base. Halpern and Vardi **?** advocated a model-checking approach. In this approach, theorem proving is replaced by evaluating queries against an appropriate semantic model of the agent's knowledge. This can be viewed as a *knowledge-level* approach to knowledge bases (**?**). Such a semantic model was in fact provided by Levesque **?**; he associates with a knowledge base a set of truth assignments. We describe here a different semantic approach. We show that an operational semantics for knowledge bases, based on the model of multi-agent systems from (**?**) (which in turn, is based on earlier models that appeared in (**?**; **?**; **?**; **?**; **?**)), offers a clean and intuitive knowledge-level model. The basic idea of this approach is to model the system as a set of possible behaviors. Knowledge is then ascribed to agents according to the possible-worlds principle: a fact φ is known to an agent a if φ holds in all the states of the system that a considers possible. Thus, in our approach knowledge "falls out" of the operational model of the system. We argue that this approach offers significant advantages compared to previous approaches to modeling knowledge bases.

Knowledge in multi-agent systems

We briefly review the framework of (**?**) for modeling multi-agent systems. The basic idea of this approach is to model systems *operationally* (in the spirit of the operational-semantics approach to programming languages (**?**; **?**)). We assume that at each point in time, each agent is in some *local state*. Informally, this local state encodes the information the agent has observed thus far. In addition, there is also an *environment* state, that keeps track of everything relevant to the system not recorded in the agents' states. The way we split up the system into agents and environment depends on the system being analyzed.

A *global state* is a tuple $(s_e, s_1, \ldots, s_n)$ consisting of the environment state s_e and the local state s_i of each agent i. A *run* of the system is a function from time (which, for ease of exposition, we assume ranges over the natural numbers) to global states. Thus, if r is a run, then $r(0), r(1), \ldots$ is a sequence of global states that, roughly speaking, is a complete description of what happens over time in one

---

*This paper appears in *AAAI-94 (Proceedings of the Twelfth National Conference on Artificial Intelligence)*, pp. 1142–1147, 1994.

†Work supported in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080.

‡Currently on sabbatical at Oxford; work supported in part by a Helen and Milton A. Kimmelman career development chair.

§This research was done while this author was at the IBM Almaden Research Center.

possible execution of the system. We take a *system* to consist of a set of runs. Intuitively, these runs describe all the possible sequences of events that could occur.

Given a system $\mathcal{R}$, we refer to a pair (r, m) consisting of a run $r \in \mathcal{R}$ and a time m as a *point*. If $r(m) = (s_e, s_1, \ldots, s_n)$, we define $r_e(m) = s_e$ and $r_i(m) = s_i$, for $i = 1, \ldots, n$; thus, $r_e(m)$ is the environment state and $r_i(m)$ is process i's local state at the point (r, m). We say that two points (r, m) and (r', m') are *indistinguishable* to agent i, and write $(r, m) \sim_i (r', m')$, if $r_i(m) = r'_i(m')$, i.e., if agent i has the same local state at both points. Finally, we define an *interpreted system* to be a pair $(\mathcal{R}, \pi)$ consisting of a system $\mathcal{R}$ and a mapping π that associates a truth assignment to the primitive propositions at each global state.

An interpreted system can be viewed as a Kripke structure: the points are the possible worlds, and $\sim_i$ plays the role of the accessibility relation. We give semantics to knowledge formulas in interpreted systems just as in Kripke structures: Given a point (r, m) in an interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$, we have $(\mathcal{I}, r, m) \models K_i\varphi$ (that is, the formula $K_i\varphi$ is satisfied at the point (r, m) of $\mathcal{I}$) if $(\mathcal{I}, r', m') \models \varphi$ for all points (r', m') such that $(r', m') \sim_i (r, m)$. Notice that under this interpretation, an agent knows φ precisely if φ is true at all the situations the system could be in, given the agent's current information (as encoded by its local state). Since $\sim_i$ is an equivalence relation, knowledge in this framework satisfies the S5 axioms.

The major application of this framework has been in providing a knowledge-level analysis of distributed protocols (**?**). It is often relatively straightforward to construct the system corresponding to a given protocol. The local state of each process can typically be characterized by a number of local variables (which, for example, describe the messages received and the values of certain local registers). The runs describe the behavior of the system as a result of running the protocol. (See (**?**) for detailed examples of the modeling process.) Here we examine how this framework can be used to model knowledge bases.

Knowledge bases as multi-agent systems

Following Levesque **?**, we view a KB as a system that is told facts about an external world, and is asked queries about that world.[1] The standard approach in AI to modeling knowledge bases is just to identify a KB with a formula, or set of formulas, that can informally be thought of as describing what the KB knows. When the KB is asked a query ψ, it computes (using some computational procedure) whether ψ holds. Levesque takes a more semantic approach, associating with the KB the set of truth assignments that the KB considers possible at any time, as a function of what it has been told.

We now show how knowledge bases can be modeled as multi-agent systems. As we shall see, doing so gives us a number of advantages. Basically, since we are modeling knowledge bases operationally, we can easily capture aspects that are hard to capture in a symbolic model or even in Levesque's knowledge-level model. For one thing, we can capture assumptions about how the KB obtains its knowledge and show how these assumptions affect the KB's knowledge. Furthermore, the model allows us to study how the KB's knowledge evolves with time. Generally, the model is very flexible and can easily be adapted to many applications.

The first step in modeling the KB in our framework is to decide who the agents are and what the role of the environment is. The KB is clearly an agent in the system. In addition, we choose to have another agent called the *Teller*; this is the agent that tells the KB facts about the external world. We use the environment to model the external world. It is possible to use the environment to also model the Teller, but, as we shall see later on, our approach offers certain advantages. We want to view the environment's state as providing a complete description of (the relevant features of) the external world, the local state of the KB as describing the information that the KB has about the external world, and the local state of the Teller as describing the information that the Teller has about the external world and about the KB. This allows us to distinguish what is true (as modeled by the environment's state) from what is known to the Teller (as modeled by the Teller's state) and from what the KB is told (as modeled by the KB's state).

That still gives us quite a bit of freedom in deciding how to model the global states. If we can describe all the relevant features of the external world by using a set Φ of primitive propositions, then we can take the environment to be just a truth assignment to the primitive propositions in Φ. If, instead, we need to use first-order information to describe the world, then we can take the environment to be a relational structure.

What about the KB's local state? We want it to represent all the relevant information that the KB has learned. We can do this by taking the local state to consist of the sequence of facts that the KB has been told and queries that it has been asked. If we assume that the sequence of queries does not carry any information about the external world, then we can simplify this representation by including in the local state only the sequence of facts that the KB has been told, and ignoring the queries. This is in fact what we do.

Finally, the Teller's state has to describe the Teller's information about the external world and about the KB. We assume that the Teller has complete information about the KB, since the Teller is the sole source for the KB's information. Thus, the Teller's local state contains a description of its information about external world as well as the sequence of facts that the KB has been told.

What does the KB know after it has been told some fact φ? Assuming that what it has been told is true, it may seem reasonable to say that the KB knows φ. This is clearly false, however, if the external world can change. It might well be the case that φ was true when the KB was told it, and is no longer true afterwards. For definiteness, we assume that the external world is stable. As we shall see, even with this assumption, if φ can include facts about the

[1] Levesque models this in terms of *TELL* and *ASK* operations.

KB's knowledge, then φ may be true when the KB is told it, but not afterwards.

To get a feel for some of the issues involved, we focus first on modeling a fairly simple concrete situation. We later consider what happens when we weaken these assumptions. We assume that:

1. the external world can be described propositionally, using the propositions in a finite set Φ,
2. the external world is stable, so that the truth values of the primitive propositions describing the world do not change over time, at least for the intervals of time we are analyzing,
3. the Teller has complete information about the external world and about the KB,
4. the KB is told and asked facts only about the external world, and not facts about its own knowledge, and these facts are expressed as propositional formulas,
5. everything the KB is told is true, and
6. there is no *a priori* initial knowledge about the external world, or about what the KB will be told.

The first assumption tells us that we can represent the environment's state as a truth assignment α to the primitive propositions in Φ. The second assumption tells us that in each run r, the environment's state $r_e(m)$ is independent of m; the environment's state does not change over time. As observed by Katsuno and Mendelzon **?**, this is the assumption that distinguishes *belief revision* from *belief update*. The third assumption tells us that the Teller's state includes the truth assignment α, which describes the external world. Given that we are representing the KB's local state as a sequence of facts that it has been told, the fourth assumption tells us that this local state has the form $\langle \varphi_1, \ldots, \varphi_k \rangle$, $k \geq 0$, where $\varphi_1, \ldots, \varphi_k$ are propositional formulas. We assume that the Teller's local state has a similar form, and consists of the truth assignment that describes the real world, together with the sequence of facts it has told the KB. Thus, we take the Teller's local state to be of the form $(\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle)$, where α is a truth assignment and $\varphi_1, \ldots, \varphi_k$ are propositional formulas. Since the Teller's state is simply the pair consisting of the environment's state and the KB's state, we do not represent it explicitly, but rather denote a global state by $(\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle, \cdot)$. The fifth assumption tells us that everything that the KB is told is true. This means that in a global state of the form $(\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle, \cdot)$, each of $\varphi_1, \ldots, \varphi_k$ must be true under truth assignment α. The part of the sixth assumption that says that there is no initial knowledge of the world is captured by assuming that the initial state of every run has the form $(\alpha, \langle\, \rangle, \cdot)$, and that for every truth assignment α', there is some run with initial global state $(\alpha', \langle\, \rangle, \cdot)$. We capture the second half of the sixth assumption—that there is no knowledge about what information will be given—by not putting any further restrictions on the set of possible runs. We discuss this in more detail later.

To summarize, we claim our assumptions are captured by the interpreted system $\mathcal{I}^{kb} = (\mathcal{R}^{kb}, \pi^{kb})$, where $\mathcal{R}^{kb}$ consists of all runs r such that for some sequence $\varphi_1, \varphi_2, \ldots$ of propositional formulas and for some truth assignment α:

- **KB1**. $r(0) = (\alpha, \langle\, \rangle, \cdot)$
- **KB2**. if $r(m) = (\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle, \cdot)$, then
  1. either $r(m+1) = r(m)$, or $r(m+1) = (\alpha, \langle \varphi_1, \ldots, \varphi_k, \varphi_{k+1} \rangle, \cdot)$,
  2. $\varphi_1 \wedge \cdots \wedge \varphi_k$ is true under truth assignment α, and
  3. $\pi^{kb}(r, m) = \alpha$, that is, π^{kb} is defined so that the truth assignment at (r, m) is given by the environment's state.

Our assumption that $\mathcal{R}$ consists of *all* runs that satisfy the conditions above also captures the assumption that there is no knowledge about what information will be given. This is perhaps best understood by example. There may be *a priori* knowledge that, if p is true, then this is the first thing the KB will be told. This places a restriction on the set of possible runs, eliminating runs with global states of the form $(\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle, \cdot)$ such that $k \geq 1$ and p is true under the truth assignment α, but $\varphi_1 \neq p$. It is easy to construct other examples of how what information is given or the order in which it is given might impart knowledge beyond the facts themselves. By allowing all runs r consistent with KB1 and KB2 in $\mathcal{R}$, we are saying that there is no such knowledge.

Having defined the system $\mathcal{I}^{kb}$, we can see how the KB answers queries. Suppose that at a point (r, m) the KB is asked a query ψ, where ψ is a propositional formula. Since the KB does not have direct access to the environment's state, ψ should be interpreted not as a question about the external world, but rather as a question about the KB's knowledge of the external world. Thus, the KB should answer "Yes" exactly if $(\mathcal{I}^{kb}, r, m) \models K_{KB}\psi$ holds (taking K_{KB} to denote "the KB knows"), "No" exactly if $(\mathcal{I}^{kb}, r, m) \models K_{KB}\neg\psi$ holds, and "I don't know" otherwise.

Suppose the KB is in local state $\langle \varphi_1, \ldots, \varphi_k \rangle$. We can view the formula $\kappa = \varphi_1 \wedge \cdots \wedge \varphi_k$ as a summary of its knowledge about the world; the KB knows only what follows from this. This could be interpreted in two ways: the KB could answer "Yes" exactly if ψ is a consequence of κ, or if $K_{KB}\psi$ is a consequence of $K_{KB}\kappa$. As the following result shows, these two interpretations are equivalent.

Proposition 1: *Suppose that $r_{KB}(m) = \langle \varphi_1, \ldots, \varphi_k \rangle$. Let $\kappa = \varphi_1 \wedge \cdots \wedge \varphi_k$ and let ψ be a propositional formula. The following are equivalent:*

(a) $(\mathcal{I}^{kb}, r, m) \models K_{KB}\psi$.

(b) $\kappa \Rightarrow \psi$ *is a propositional tautology.*

(c) $K_{KB}\kappa \Rightarrow K_{KB}\psi$ *is a valid formula in S5.*

Thus, Proposition **??** shows that under our assumptions, we can model the KB in the standard AI manner: as a formula. Moreover, in order to answer a query, the KB must compute what follows from the formula that represents its knowledge.

Proposition **??** characterizes how the KB answers propositional queries. As argued by Levesque **?** and Reiter **?**, in general the KB may have to answer non-propositional queries. How should the KB handle such queries as $(p \Rightarrow K_{KB}p)$ ("if p is the case, then the KB knows that it is the case")? Here also we want the KB to answer "Yes" to a query φ exactly if $(\mathcal{I}^{kb}, r, m) \models K_{KB}\varphi$, "No" exactly if $(\mathcal{I}^{kb}, r, m) \models K_{KB}\neg\varphi$ holds, and "I don't know" otherwise. When does the formula $K_{KB}(p \Rightarrow K_{KB}p)$ hold? It is not hard to show that this formula is equivalent to $K_{KB}p \vee K_{KB}\neg p$, so the answer to this query already follows from Proposition **??**: the answer is "Yes" if either p follows from what the KB has been told, or $\neg p$ does, and "I don't know" otherwise. It is not possible here for the answer to be "No", since $K_{KB}\neg(p \Rightarrow K_{KB}p)$ is equivalent to $K_{KB}(p \wedge \neg K_{KB}p)$, which is easily seen to be inconsistent with S5.

We are mainly interested in what can be said about formulas that involve only the KB's knowledge, since we view the Teller as being in the background here. We define a *KB-formula* to be one in which the only modal operator is K_{KB}; a *KB-query* is a query which is a KB-formula. Standard arguments from modal logic can be used to show that for every KB-formula of the form $K_{KB}\varphi$ we can effectively find an equivalent formula that is a Boolean combination of formulas of the form $K_{KB}\psi$, where ψ is propositional. It follows that the way that the KB responds to KB-queries can already be determined from how it responds to propositional queries. The reason is as follows. To decide on its answer to the query φ, we must determine whether $K_{KB}\varphi$ holds and whether $K_{KB}\neg\varphi$ holds. As we just noted, we can effectively find a formula equivalent to $K_{KB}\varphi$ that is a Boolean combination of formulas of the form $K_{KB}\psi$, where ψ is propositional, and similarly for $K_{KB}\neg\varphi$. We then need only evaluate formulas of the form $K_{KB}\psi$, where ψ is propositional. Thus, using Proposition **??**, we can compute how the KB will answer KB-queries from the conjunction of the formulas that the KB has been told.

There is another way of characterizing how the KB will answer KB-queries. Given a propositional formula φ, let S^{φ} consist of all truth assignments α to propositions in Φ such that φ is true under truth assignment α. Let $M^{\varphi} = (S^{\varphi}, \pi, \mathcal{U})$ be the Kripke structure such that $\pi(\alpha) = \alpha$ and $\mathcal{U}$ is the universal relation (so that for all $\alpha, \beta \in S^{\varphi}$, we have $(\alpha, \beta) \in \mathcal{U}$). In a sense, we can think of M^{φ} as a *maximal* model of φ, since all truth assignments consistent with φ appear in M^{φ}. As the following result shows, if κ is the conjunction of the formulas that the KB has been told, then for an arbitrary formula ψ, the KB knows ψ exactly if $K_{KB}\psi$ holds in the maximal model for κ. Intuitively, if the KB was told κ, then *all* that the KB knows is κ. The maximal model for κ is the model that captures the fact that κ is all that the KB knows.

Proposition 2: *Suppose that* $r_{KB}(m) = \langle \varphi_1, \ldots, \varphi_k \rangle$*, and* $\kappa = \varphi_1 \wedge \cdots \wedge \varphi_k$*. Then for all KB-formulas* ψ*, we have that* $(\mathcal{I}^{kb}, r, m) \models \psi$ *iff* $(M^{\kappa}, r_e(m)) \models \psi$.

Levesque **?** defines M^{κ} as the knowledge-level model of the KB after it has been told $\varphi_1, \ldots, \varphi_k$. Thus, Proposition **??** shows that in the propositional case, our operational model is equivalent to Levesque's knowledge-level model.

Our discussion so far illustrates that it is possible to model a standard type of knowledge base within our framework. But what do we gain by doing so? For one thing, it makes explicit the assumptions underlying the standard representation. In addition, we can talk about what the KB knows regarding its knowledge, as shown in Proposition **??**. Beyond that, as we now show, it allows us to capture in a straightforward way some variants of these assumptions. The flexibility of the model makes it easier to deal with issues that arise when we modify the assumptions.

We begin by considering situations where there is some prior knowledge about what information will be given. As we observed earlier, the fact that we consider *all* runs in which KB1 and KB2 are true captures the assumption that no such knowledge is available. But, in practice, there may well be default assumptions that are encoded in the conventions by which information is imparted. We earlier gave an example of a situation where there is a convention that if p is true, then the KB will be told p first. Such a convention is easy to model in our framework: it simply entails a restriction on the set of runs in the system. Namely, the restriction is that for every point (r, m) in the system where $r(m) = (\alpha, \langle \varphi_1 \rangle, \cdot)$, we have $\varphi_1 = p$ iff p is true under α. Recall that the order in which the KB is given information is part of its local state. In a precise sense, therefore, the KB knows what this order is. In particular, it is straightforward to show that, given the above restriction, the KB either knows p or knows $\neg p$ once it has been told at least one fact.

In a similar fashion, it is easy to capture the situation where there is some *a priori* knowledge about the world, by modifying the set of runs in $\mathcal{I}^{kb}$ appropriately. Suppose, for example, that it is known that the primitive proposition p must be true. In this case, we consider only runs r such that $r_e(0) = \alpha$ for some truth assignment α that makes p true. An analogue to Proposition **??** holds: now the KB will know everything that follows from p and what it has been told.

Next, consider the situation where the Teller does not have complete information about the world (but still has complete information about the KB). We model this by including in the Teller's state a nonempty set $\mathcal{T}$ of truth assignments. Intuitively, $\mathcal{T}$ is the set of possible external worlds that the Teller considers possible. The set $\mathcal{T}$ replaces the single truth assignment that describes the actual external world. Since we are focusing on knowledge here, we require that $\alpha \in \mathcal{T}$; this means that the true external world is one of the Teller's possibilities. The Teller's state also includes the sequence of facts that the KB has been told. To avoid redundancy, we denote the Teller's state by $\langle \mathcal{T}, \cdot \rangle$. Global states now have the form $(\alpha, \langle \varphi_1, \ldots, \varphi_k \rangle, \langle \mathcal{T}, \cdot \rangle)$. We still require that everything the KB is told be true; this means that the Teller tells the KB "φ" only if φ is true in all the truth assignments in $\mathcal{T}$. It is easy to see that this means that the Teller says φ only if $K_T\varphi$ holds (taking K_T to denote "the Teller knows"). Not surprisingly, Propositions **??** and **??** continue to hold in this setting, with essentially no change in proof.

Once we allow the Teller to have a collection $\mathcal{T}$ of worlds that it considers possible, it is but a short step to allow the Teller to have false beliefs, which amounts to allowing $\mathcal{T}$ not to include the actual world. We would still require that the Teller tells the KB φ only if φ is true in all the truth assignments in $\mathcal{T}$. In this case, however, this means that the Teller only *believes* φ to be the case; its beliefs may be wrong. How should we ascribe beliefs to agents in a multi-agent system? In the scenario described here, the KB and the Teller believe that the Teller is truthful, so they both consider some global states to be impossible, namely, the global states in which $\alpha \notin \mathcal{T}$. Thus, it makes sense here to change the definition of the accessibility relation in the Kripke structure associated with a system in order to make global states where $\alpha \notin \mathcal{T}$ inaccessible. The possible-worlds principle now ascribes beliefs rather than knowledge; see (**?**) for details.[2]

Knowledge-based programs

Up to now we have assumed that the KB is told only propositional facts. Things get somewhat more complicated if the KB is given information that is not purely propositional; this in fact is the situation considered by Levesque **?**. For example, suppose the KB is told $p \Rightarrow K_{KB}p$. This says that if p is true, then the KB knows it. Such information can be quite useful, assuming that the KB can actually check what it knows and does not know. In this case, the KB can check if it knows p; if it does not, it can then conclude that p is false. As this example shows, once we allow the KB to be given information that relates its knowledge to the external world, then it may be able to use its introspective abilities to draw conclusions about the external world.

It is now not so obvious how to represent the KB's knowledge symbolically, i.e., by a formula. One complication that arises once we allow non-propositional information is that we can no longer assume that the KB knows everything it has been told. For example, suppose the primitive proposition p is true of the external world, and the KB has not been given any initial information. In this situation, the formula $p \wedge \neg K_{KB}p$ is certainly true. But after the KB is told this, then it is certainly not the case that the KB knows $p \wedge \neg K_{KB}p$; indeed, as we noted earlier, $K_{KB}(p \wedge \neg K_{KB}p)$ is inconsistent with S5. Nevertheless, the KB certainly learns something as a result of being told this fact: it learns that p is true. As a result, $K_{KB}p$ should hold after the KB is told $p \wedge \neg K_{KB}p$. Thus, we cannot represent the KB's knowledge simply by the conjunction of facts that it has been told, even if they are all true.

Levesque **?** describes a knowledge-level model for the KB's knowledge in this case. After the KB has been told the sequence $\varphi_1, \ldots, \varphi_k$, it is modeled by a Kripke structure $M^{\varphi_1,\ldots,\varphi_k}$, which we define inductively. The initial model is $M^\epsilon = (S^\epsilon, \pi, \mathcal{U})$, where S^ϵ is the set of all truth assignment to the propositions in Φ, $\pi(\alpha) = \alpha$, and $\mathcal{U}$ is the universal relation. Suppose that $M^{\varphi_1,\ldots,\varphi_{k-1}} = (S^{\varphi_1,\ldots,\varphi_{k-1}}, \pi, \mathcal{U})$ has been defined. Then $M^{\varphi_1,\ldots,\varphi_k} = (S^{\varphi_1,\ldots,\varphi_k}, \pi, \mathcal{U})$, where $S^{\varphi_1,\ldots,\varphi_k} = \{w \in S^{\varphi_1,\ldots,\varphi_{k-1}} \mid (M^{\varphi_1,\ldots,\varphi_{k-1}}, w) \models \varphi_k\}$. As in the earlier discussion of the maximal model, this definition attempts to capture the idea that the KB knows only what it has been told. The induction construction ensures that this principle is applied whenever the KB is told a formula.

In our approach, we need to be able to describe the system that results when the KB is given information that may involve its own knowledge. As before, we take the KB's local state to consist of a sequence of formulas, except that we now allow the formulas to be modal formulas which can talk about the KB's knowledge, not just propositional formulas. The only difficulty comes in restricting to runs in which the KB is told only true formulas. Since we are now interested in formulas that involve knowledge, it is not clear that we can decide whether a given formula is true without having the whole system in hand. But our problem is to construct the system in the first place!

While it is difficult to come up with an explicit description of the system, it is easy to describe this system implicitly. After all, the behavior of the agents here is fairly simple. The Teller here can be thought as following a *knowledge-based program* (**?**; **?**; **?**). This is a program with explicit tests for knowledge. Roughly speaking, we can think of the Teller as running a nondeterministic program TELL that has an infinite collection of clauses, one for each formula φ, of the form:

if $K_T\varphi$ **do** $\mathsf{tell}(\varphi)$.

Intuitively, when running this program, the Teller nondeterministically chooses a formula φ that it knows to be true, and tells the KB about it. The propositional case considered in the previous section corresponds to the Teller running the analogous knowledge-based program $\mathsf{TELLPROP}$ in which the formulas φ are restricted to be propositional. Using techniques introduced in (**?**), it can be shown that both TELL and $\mathsf{TELLPROP}$ can be associated with unique interpreted systems $\mathcal{I}^{tell}$ and $\mathcal{I}^{tellprop}$, respectively. It turns out that the interpreted system $\mathcal{I}^{kb}$ (defined in the previous section), which captured the interaction of the KB with the Teller in the propositional case, is precisely the system $\mathcal{I}^{tellprop}$. This observations provides support for our intuition that $\mathcal{I}^{tell}$ appropriately captures the situation where the Teller tells the KB formulas that may involve the KB's knowledge. Moreover, the system that we get is closely related to Levesque's knowledge-level model described earlier.

Proposition 3: *Suppose that* $r_{KB}(m) = \langle\varphi_1, \ldots, \varphi_k\rangle$. *Then for all KB-formulas* ψ, *we have that* $(\mathcal{I}^{kb}, r, m) \models \psi$ *iff* $(M^{\varphi_1,\ldots,\varphi_k}, r_e(m)) \models \psi$.

One advantage of using knowledge-based programs is that we can consider more complicated applications. In many such applications, one cannot divide the world neatly into a KB and a Teller. Rather, one often has many agents, each of which plays both the role of the KB and the Teller. For example, suppose that we have n agents, each of whom makes an initial observation of the external world and then communicates with the others. We assume that the agents

[2] See also (**?**) for a general approach to adding belief to this framework.

are truthful, but that they do not necessarily know or tell the "whole truth". We can view all the agents as following knowledge-based programs similar to TELL. At every round, agent i nondeterministically selects, for each agent j, a formula φ_j that i knows to be true, and "tells" φ_j to j. Formally, agent i's program consists of all clauses of the form:

if $K_i\varphi_1 \wedge \ldots \wedge K_i\varphi_k$ **do** $\mathsf{send}(\varphi_1, j_1)$;...;$\mathsf{send}(\varphi_k, j_k)$,

where we take $\mathsf{send}(\varphi_l, j_l)$ to be the action sending the message φ_l to agent j_l. Here we allow the messages φ to be arbitrary modal formulas; for example, Alice can tell Bob that she does not know whether Charlie knows a fact p.

In this case, it is no longer clear how to model the agents symbolically or at the knowledge level as in (**?**).[3] In fact, while it is easy to characterize the appropriate interpreted system implicitly, via knowledge-based programs, it is quite difficult to describe the system explicitly. Nevertheless, our approach enables us to characterize the agents' knowledge in this case and analyze how it evolves with time. We view this as strong evidence to the superiority of the operational approach.

Conclusions

We have tried to demonstrate the power of the operational approach to modeling knowledge bases. We have shown that under simple and natural assumptions, the operational approach gives the same answers to queries as the more standard symbolic approach and Levesque's knowledge-level approach. The advantage of the operational approach is its flexibility and versatility. We have given some evidence of this here. Further evidence is provided by the recent use of this framework (extended to deal with beliefs) to model belief revision and belief update (**?**; **?**). We are confident that the approach will find yet other applications.

[3] See (**?**) for a general framework for mutual belief revision.