Grounding Knowledge and Action Selection in Agent-Based Systems

Natasha Alechina   Brian Logan

School of Computer Science and IT
University of Nottingham
Nottingham NG8 1BB, UK
{nza,bsl}@cs.nott.ac.uk

Abstract

We show how to modify the model of knowledge based on interpreted systems developed by Halpern and Moses to give a new class of interpreted systems, *representational interpreted systems*, in which the fact that an agent knows that ϕ corresponds to a particular configuration in the agent's internal state. This permits an objective and constructive test for the agent's knowledge in a way similar to the situated automata approach of Rosenschein and Kaelbling. We prove that representational interpreted systems have the same logical theory as interpreted systems, i.e., they satisfy the same formulas of epistemic logic, and show how our approach can be used to ground action selection in the state of the agent, allowing us to verify constructively that the selection of a particular action in a particular environment is correct.

1 Introduction

Formal models of agent systems are central both to deepening our understanding of the notion of agency and to the principled design of agent systems. The concept of knowledge plays a key role in many formal models in providing an abstract account of the agent's state and behaviour. Epistemic notions such as knowledge and belief provide a compact and powerful way of describing, explaining and predicting the behaviour of agents.

A common approach is to model the agent in an epistemic logic and prove theorems about the agent's behaviour in that logic (see, for example, [10, 4, 9, 13, 14, 5, 12, 19, 18, 21]). Such an approach is useful as an abstraction tool even when we have perfect knowledge of the design of the system, but can also be applied when the system in question is not known to or is known not to employ intentional notions[1]. In this paper we are concerned with ascribing knowledge to agents. Such knowledge is often termed implicit knowledge and includes as a subset the agent's explicit knowledge, i.e., the knowledge the agent has access to and can answer questions about (as studied in e.g., [7]). However if we are ascribing knowledge we then have to define precisely what we mean by 'the agent knows that ϕ'. In many cases, the semantics of agent states is simply stipulated by the designer, who assigns a particular meaning to a particular state or set of states, and strives to design the agent in such a way that the stipulated semantics holds. However such stipulations, by their very nature, cannot be objective, and are prone to over-interpretation [11]. A key component of any theory of agent systems is a formal model of agent knowledge which makes precise what it is for an agent to know that ϕ. Such a model should be *computationally grounded* (in the sense of [20]) and *constructive*. By computationally grounded we mean that knowledge should be interpreted in terms of concrete computational notions such as the runs of a program or system. By constructive we mean that we can tell whether an agent knows that ϕ solely by inspecting the properties of the agent (typically the configuration of the agent's state). In addition, the model should not require unreasonable properties of the agent (e.g., that the agent's state is infinite).

In this paper, we present a formal model of knowledge ascription for agents with partial knowledge which permits an objective and constructive test for the agent's knowledge. In the next section we introduce our theoretical framework and formulate some postulates for a general model of knowledge ascription. We then consider two existing models of knowledge in the light of these criteria, the objective correlation approach of Rosenschein and Kaelbling [16] and the interpreted systems approach of Halpern and Moses [6]. In section 5 we then show how to combine the interpreted systems and objective correlation approaches to give a new class of interpreted systems, *representational interpreted systems*, in which the fact that an agent knows that ϕ corresponds to a particular configuration in the agent's internal state. We prove that representational interpreted systems have the same logical theory as interpreted systems, i.e., they satisfy the same formulas of epistemic logic. Finally, in section 6 we show how our approach can be used to ground action selection in the state of the agent, allowing us to verify constructively that the selection of a particular action in a particular environment is correct.

[1] Dennett [3] has termed this latter strategy "adopting the intentional stance."

2 Background

We begin by stating some background assumptions which should be satisfied by any model of knowledge ascription applicable to agents with partial knowledge. Our model of agents is deliberately abstract. We assume only that an agent has a *state* which can be viewed as a set of variables or *locations* (following the terminology of [16]) together with the sets of values which can be assumed by the locations. The set of possible states of the agent is the cross product of sets of values for each location.

We assume that our agents and the implicit knowledge which we ascribe to them satisfy the following postulates:

1. the agent may have only partial knowledge of its environment;
2. the agent can add to its knowledge over time;
3. the agent's knowledge is uniquely encoded in its state; and
4. the agent's state is finite.

We do not claim that these postulates exhaust the desirable properties of a theory of knowledge ascription, rather our aim is to establish a minimal framework for ascribing knowledge. For example, a necessary property when ascribing knowledge to imperfect reasoners is the absence of logical omniscience (see, for example, [1, 2]); however implicit knowledge can be safely assumed to be closed under logical consequence (see, for example, [10, 17, 5]).

The first postulate is necessary to model agents which are resource bounded in the most general sense, for example, agents whose sensors do not return complete knowledge of the environment or themselves. In particular, we do not require that the agent know all facts which are true of the environment—it should be possible that $\phi \wedge \neg K_i\phi$, where $K_i\phi$ means that agent i knows that ϕ. The second postulate allows the agent to acquire new knowledge, (up to the point at which its representational capacity is exhausted, see below). For example, if the agent looks out of the window it may discover that it is raining. The third postulate requires that the agent's knowledge is encoded in some properties of its state. This guarantees that our model of knowledge ascription is objective and constructively verifiable. The fourth postulate is simply a restriction on all practical agents. Taken together, the third and fourth postulates imply that the agent can only know finitely many non-equivalent formulas.

2.1 Propositional epistemic logic

We model an agent's implicit knowledge using propositional epistemic logic. A language of propositional epistemic logic contains a set of propositional variables (primitive propositions) $\Phi = \{p_1, \ldots, p_n, \ldots\}$, boolean connectives $\neg, \wedge$ (and other definable connectives) and unary modal operators K_i for every agent i in the set of agents $\mathcal{A} = \{1, \ldots, m\}$. A well formed formula (w.f.f.) is defined as follows: if p is in Φ, it is a w.f.f.; if ϕ and ψ are well formed formulas, then so are $\neg\phi$ and $\phi \wedge \psi$; if ϕ is a w.f.f. then so is $K_i\phi$ (where $i \in \mathcal{A}$) which stands for 'agent i knows that ϕ'. Nothing else is a w.f.f.

The language of epistemic logic is usually interpreted with respect to (multi-modal) Kripke structures. A Kripke structure $M = (W, \pi, \{R_i : i \in \mathcal{A}\})$ is a triple where W is a non-empty set of possible worlds (points, states), π is a function which for every possible world and every propositional variable in Φ produces a truth value of that proposition in that world ($\mathbf{true}$ or $\mathbf{false}$), and for every $i \in \mathcal{A}$, R_i is a binary relation on W called the accessibility relation. The relation $M, w \models \phi$ where $w \in W$ (ϕ is true in w) is defined inductively as follows:

For a propositional variable p, $M, w \models p$ iff $\pi(w, p) = \mathbf{true}$.

$M, w \models \neg\psi$ iff $M, w \not\models \psi$.

$M, w \models \psi \wedge \psi'$ iff $M, w \models \psi$ and $M, w \models \psi'$.

$M, w \models K_i\psi$ iff for all w' such that $R_i(w, w')$, $M, w' \models \psi$.

For the logic of knowledge, R_i is usually assumed to be an equivalence relation on W. The resulting logic is called S5.

In what follows we will use the language of epistemic logic to formulate the properties of knowledge arising from different computational interpretations of knowledge, and follow [5] considering Kripke structures corresponding to those interpretations.

3 Knowledge as Objective Correlation

Perhaps the most straightforward approach to ascribing knowledge to an agent is to define knowledge as an objective correlation between the state of the agent and the state of the environment. The agent knows that ϕ if the agent's state carries information that ϕ, that is, if there is a part of the agent's state which has specific values if ϕ holds in the environment. We associate some location (or locations), x, in the agent's state with a proposition ϕ, and require that whenever x takes a particular value (or values), say 1, ϕ is true. This definition has the advantage of being objective and is constructively verifiable when the number of states in the environment is finite.

This is essentially the approach taken by Rosenschein and Kaelbling [15, 16] in their work on situated automata: "We say that when an agent ... is in state v, it *carries the information that* ϕ if and only if whenever it is in state v, ϕ is true in the world" [16]. Rosenschein and Kaelbling view an agent as an automaton made up of parts or locations, each of which can take a finite set of values. Information can either be encoded 'by location', i.e., a particular location is devoted to the representation of a particular proposition, or 'by value', i.e., the values of the locations involved in representing the locations are critical, but the same combinations of values occurring at any (sequence of) locations would carry the same information. In what follows, and without loss of generality, we only consider encoding by location. A location x having a certain value carries information about any proposition which is true whenever x has that value.

The resulting model of agent knowledge satisfies S5 and has the advantage that there is a precise relationship between the internal states of the agent and conditions in its environment, allowing an objective attribution of semantics to a value (or values) at a particular location (or set of locations). For example, it gives a precise account of the degradation of information over time: "a location containing a value at time t and continuing to hold that value until time $t + k$ will be assigned as its information content the disjunction of world conditions satisfied at any time instant in the interval $[t, t + k]$" [16].

4 Interpreted Systems

More recently, attention has switched to computationally grounded approaches inspired by ideas in philosophical logic. In philosophical logic, knowledge and belief are defined in terms of possible worlds: *a person knows/believes that* ϕ *if* ϕ *is true in all states of the world which the person considers possible* (the general idea was introduced in [8]). A natural adaptation to agents and computations is to say that an agent knows that ϕ if in all states of the system where the agent has the same local state, ϕ holds. The agent has no means of distinguishing different global states if its own local state is the same in all of them, so it considers all of them 'possible'. A well known formalisation of this approach using the notion of *interpreted systems* was given by Halpern and Moses [6] and further developed in [5]. However, while interpreted systems are grounded, they are not constructive, since they refer to a potentially infinite number of states.

One way to fix this would be to ground knowledge ascription in agent's state as in Rosenschein and Kaelbling. However, in interpreted systems, the definition of knowledge does not refer to the structure of the agent's state. In fact at different points in time the agent may be in completely different states while knowing that ϕ. The mapping from states to propositions is many-to-many, and we cannot ascribe knowledge to the agent by inspecting the agent's state. In what follows we show how to combine the interpreted systems and objective correlation approaches to give representational interpreted systems, in which the agent's knowledge corresponds to a particular configuration of the agent's internal state. We show how to modify the interpreted systems approach so that there is a precise correspondence between the contents of the agent's state and the agent's knowledge as defined in terms of all possible runs. We prove that our modification leaves the logical properties of the approach (the set of formulas of the epistemic logic which are satisfiable according to it) unaltered.

First we introduce notation and definitions based on [5].

4.1 Interpreted systems and Kripke structures

A *global state* g (of a system) consists of the state of the environment g_e and local states of all m agents in the environment, $g_1, \ldots, g_m$. A *run* is a function r from the set of time points (assumed to be natural numbers) to the set of all global states. A *system* $\mathcal{R}$ is the set of all possible runs. Denote the set of all global states occurring in a run in $\mathcal{R}$ by $\mathcal{G}(\mathcal{R})$. Let Φ be the set of primitive propositions which we want to use when talking about the system. An *interpreted system* $\mathcal{I}$ is a system $\mathcal{R}$ together with a function π which for every propositional letter $p \in \Phi$ and every global state $g \in \mathcal{G}(\mathcal{R})$ gives a truth value of p in g.

Every interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$ has a corresponding Kripke structure $M = (W, \pi^M, \{R_1, \ldots, R_m\})$ where

W is the set of all pairs (r, n) where $r \in \mathcal{R}$ and n is a natural number (time point);

$\pi^M((r, n), p) = \pi(r(n), p)$;

(r, n) and (r', n') are in the relation R_i if in the global states $r(n)$ and $r'(n')$ the agent i is in the same local state.

Observe that in the absence of temporal operators which can use the information provided by time points, the truth of a formula at (r, n) really depends only on the global state $r(n)$. In what follows we make use of the lemma:

Lemma 1 *Let $\mathcal{I} = (\mathcal{R}, \pi)$ be an interpreted system, M a corresponding Kripke structure as defined above, and $M' = (W', \pi', \{R'_1, \ldots, R'_m\})$ a different Kripke structure defined as follows:*

$W' = \mathcal{G}(\mathcal{R})$;

$\pi' = \pi$;

g and g' are in the relation R'_i if $g_i = g'_i$.

Then M and M' satisfy the same formulas of epistemic propositional logic: $M, (r, n) \models \phi \Leftrightarrow M', r(n) \models \phi$.

Proof. The proof consists of a simple induction on the construction of ϕ. □

Discussion In the approach based on interpreted systems, agent i knows that ϕ in a global state g if in all possible runs of the system where it has the same local state g_i, ϕ is true. This is not a constructive definition unless the system has finitely many bounded runs. We cannot establish whether the agent knows that ϕ by examining its state alone (without examining all possible runs) because there is no particular state or partial state associated with knowing that ϕ. It is easy to give an example when an agent knows that ϕ while being in several completely different states.

The main question we are addressing in this paper is whether this is inevitable or is it possible to keep the same framework but impose tighter constraints on the agent state, associating a fixed part of the state with 'knowing that ϕ' as in the Rosenschein and Kaelbling approach. In the next section, we show that it is possible to modify the interpreted systems approach by specifying in detail the representational state of the agent.

5 Encoding Knowledge in the Agent's State

Consider an interpreted system $\mathcal{I}$. Instead of abstracting from the content of the agent's state, we now want to see it as a collection of locations each holding information concerning the agent's knowledge of a particular formula. This immediately forces us to make the following assumption:

Assumption 1 Φ is finite.

In the case of a purely propositional language or even a modal language with just one S5 modality this gives finitely many non-equivalent logical formulas. If we have more than one epistemic modal operator as is the case for multi-agent modal logics, the assumption that Φ is finite no longer guarantees that we have finitely many equivalence classes. In fact, with just two modalities and one propositional variable we can generate infinitely many non-equivalent formulas: $K_1 p, K_2 K_1 p, K_1 K_2 K_1 p, \ldots$. We could declare some arbitrary finite set of formulas to be of special interest to the agent, for example, we could restrict the depth of nestings of modal operators to be at most 2. However, we believe that a better approach is to relate the set of formulas the agents may know with the set of conditions which are relevant for the agent's action selection function.

Suppose the agent can perform a finite set of actions $ACT = \{a_1, \ldots, a_x\}$ including an empty `no-op` action. When designing the agent we can introduce the following equivalence relation $=_a$ on the set of formulas of the logical language we use: $\phi =_a \psi$ if the action which we want the agent to perform in situations where ϕ is true is the same as the action we want the agent to perform when ψ is true. This gives rise to finitely many equivalence classes of formulas. We pick a representative from each class and in this way get a finite set of formulas corresponding to locations in the agent's state. Note that all logically equivalent formulas are in the same class by construction.

We can therefore assume that an agent can only have knowledge of finitely many non-equivalent formulas, $Form(\Phi) = \{\phi_1, \ldots, \phi_k\}$. We assume that the local state of an agent is a collection of locations $\{x_{\phi_1}, \ldots, x_{\phi_k}\}$ where x_{ϕ_j} (for each j with $1 \leq j \leq k$) holds a specific value corresponding to knowing that ϕ_j if and only if the agent knows that ϕ_j.

Let us use a shorthand of $x_{\phi_j} = \top$ to denote this fact, keeping in mind that x_{ϕ_j} could be a complex location and $\top$ could in fact be a combination of particular values. We call this modification of an interpreted system a *representational interpreted system*.

We are going to show that for every interpreted system there is a representational interpreted system satisfying the same formulas of epistemic logic. Before defining such a system we briefly explain the idea. We take an interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$ and a corresponding Kripke structure M and for every global state g in $\mathcal{G}(\mathcal{R})$ we look at the set of formulas the agent knows at $g : \{\phi_j : M, g \models K\phi_j\}$. Note that we talk about $M, g \models \phi$ rather than $M, (r, n) \models \phi$ but Lemma 1 shows that this is well defined since the truth of a formula only depends on the global state. Then we replace the agent's state g_i in g by a collection of locations $\{x_{\phi_1}, \ldots, x_{\phi_k}\}$ where for all j, $x_{\phi_j} = \top$ if $M, g \models K\phi_j$ and some different value $\perp$ otherwise. The new agent state contains an explicit encoding of the agent's knowledge at g and will be the same in all global states when the agent knows the same formulas.

First we need one more assumption:

Assumption 2 Let $\mathcal{I} = (\mathcal{R}, \pi)$ be an interpreted system. For every $p \in \Phi$ and $g \in \mathcal{G}(\mathcal{R})$ the value of $\pi(g, p)$ depends only on the environment state g_e.

This assumption is a departure from the definition of interpreted systems, however we believe that it is not a very drastic one. Intuitively, we assume that the propositions which talk about such properties of the agent as its velocity, speed or battery level can be easily made dependent on the environment state by making the corresponding variables be part of the environment rather than part of the agent. In this view, all that really "belongs" to the agent is its local state which we describe in terms of the agent's knowledge rather than in terms of primitive propositions.

Definition 1 *Let* $M = (W, \pi, \{R_1, \ldots, R_m\})$ *be a Kripke model corresponding to an interpreted system* $\mathcal{I}$*.* $M^r = (W^r, \pi^r, \{R^r_1, \ldots, R^r_m\})$ *is a* representational model *corresponding to* M *if it is obtained from* M *as follows:*

$W^r = \{(g_e, r(g_1), \ldots, r(g_m)) : (g_e, g_1, \ldots, g_m) \in \mathcal{G}(\mathcal{R})\}$ *where* r *is a function which takes an agent state* g_i*) and returns a new representational state containing locations* $\{x_{\phi_1}, \ldots, x_{\phi_k}\}$ *and for all* j *with* $1 \leq j \leq k$*,* $x_{\phi_j} = \top$ *if* $M, g \models K\phi_j$ *and* $\perp$ *otherwise.*

$\pi^r((g_e, -), p) = \pi((g_e, -), p)$ *(note that by Assumption 2* π *only depends on* g_e*).*

$R^r_i(g, g')$ *if in* g *and* g' *agent* i *has the same (representational) local state.*

Theorem 1 *For every formula* $\phi \in Form(\Phi)$*, a Kripke model* M *generated by an interpreted system satisfying assumptions 1 and 2, and a representational Kripke structure* M^r *corresponding to* M*,*

$$M, (g_e, g_1, \ldots, g_m) \models \phi \ \Leftrightarrow \ M^r, (g_e, r(g_1), \ldots, r(g_m)) \models \phi$$

Proof. The proof is by induction on the construction of ϕ.

If ϕ is a propositional variable, its truth value depends only on g_e: $M, (g_e, g_1, \ldots, g_m) \models p$ iff $\pi((g_e, g_1, \ldots, g_m), p) = \mathbf{true}$ iff $\pi^r((g_e, r(g_1), \ldots, r(g_m)), p) = \mathbf{true}$ iff $M^r, (g_e, r(g_1), \ldots, r(g_m)) \models p$.

The cases of $\phi = \neg\psi$ and $\phi = \psi \wedge \psi'$ are trivial;

Suppose $\phi = K\psi$.
$M, (g_e, g_1, \ldots, g_m) \models K_i\psi$ iff
for all g'_e, g'_j with $j \neq i$,
$M, (g'_e, g'_1, \ldots, g_i, \ldots, g'_m) \models \psi$
which by the induction hypothesis is if and only if for all g'_e,
$M^r, (g'_e, r(g'_1), \ldots, r(g_i), \ldots, r(g'_m)) \models \psi$ iff
$M^r, (g_e, r(g_1), \ldots, r(g_i), \ldots, r(g_m)) \models K_i\psi$.

□

Theorem 1 shows that given assumptions 1 and 2, interpreted systems and representational interpreted systems have the same logical theory.

6 Grounding Action Selection in the Agent's State

In the interpreted systems approach, there is no requirement that the knowledge ascribed to an agent is grounded in the state of the agent. While this view is appropriate for many applications, in general agents must act on their knowledge. In the foregoing, we have shown how we can ground the implicit knowledge we ascribe to an agent in the agent's state. In this section, we show how our approach can be extended to ground action selection in the state of an agent.

We begin by defining a notion of correctness for agent actions. An action is *correct* if it is at least as good as any other action, given the current state of the environment. The notion of "at least as good as" could be defined in terms of an explicit utility measure for environment states or in terms of those actions which are rational given the agent's goals or in some other way. In particular, we do not require that there be a single correct action in a given state of the environment. We shall say that a agent's action selection function is *locally correct* if it selects a correct action given the agent's knowledge. That is, if action a_ϕ is a correct action when the environment is in a state such that ϕ is true, then an action selection function is locally correct if it returns a_ϕ whenever $K\phi$. We say an action selection function is *globally correct* if it returns a_ϕ whenever ϕ is true. Global correctness implies an agent has perfect knowledge of its environment.

If we model an agent as an interpreted system, the agent's knowledge is not correlated with any particular features of the agent's state, or even with any particular state. This makes it impossible to ground action selection in the state of the agent. Different agent states may correspond to knowledge of the same proposition, and hence can't be used a precondition for action selection. However representational interpreted systems are implementable in a way which standard interpreted systems are not. In our approach, knowledge of ϕ is explicitly grounded in the state of the agent. Locations are correlated with (possibly infinitely many non equivalent) formulas. If the selection of an action is conditional on knowledge that ϕ (or some equivalence class of propositions), this in turn means that the action should be selected when a certain unique location (or set of locations) has a particular value. We can use this fact to write a rule or procedure (or induce a finite state machine as in [16]) which looks at the appropriate locations in the agent's state and triggers a correct action. Given an appropriate choice of $=_a$ equivalence and an appropriate encoding between locations and sets of $=_a$ equivalent formulas, we can require that locations correspond to the weakest preconditions for actions.

7 Conclusion

Many researchers (e.g., [20]) have argued that models of agent knowledge should be computationally grounded, and several computationally grounded definitions of knowledge have been proposed in the literature. We believe that it is also important that models of knowledge be constructive in the sense that is possible to verify whether an agent has a knowledge of some proposition ϕ by inspecting the agent's state. Many of the computationally grounded models of knowledge are not constructive in this sense. For example, while the influential definition of knowledge based on the notion of interpreted systems [5] provides a definition of knowledge based on computational notions, it abstracts completely from the agent's state. Under this definition, an agent knows that ϕ in a global state g if in all possible runs of the system in all global states where the agent has the same internal state as in g, ϕ is true. In most practical systems it is impossible to inspect all possible runs and to verify whether the agent knows that ϕ in a given state.

In this paper we show that is possible to assume that some set of variables or locations in the agent's state always have particular values when the agent knows that ϕ (a property which can be verified constructively) while retaining the definition of knowledge based on the interpreted systems. While knowing that ϕ does not correspond to a particular part of the agent's state in the original interpreted system, it is possible to define a variant of interpreted systems called representational interpreted systems where this is the case. We show that this restriction does not change the axiomatic properties of knowledge, namely that the logical theory of representational interpreted systems and the logical theory of interpreted systems are the same. We then show how representational interpreted systems can be used to ground action selection in the state of agent, allowing us to verify constructively that the selection of a particular action in a particular environment is correct.

Our ultimate goal is a formal theory of knowledge which allows to specify agents which select appropriate actions given their knowledge. We believe that by explicitly grounding the ascription of knowledge and the selection of correct actions in the state of the agent, we have taken the first step in a design methodology that is similar in spirit to that of Rosenschein and Kaelbling.

References

[1] Natasha Alechina and Brian Logan. Logical omniscience and the cost of deliberation. In R. Nieuwenhuis and A. Voronkov, editors, *Proceedings of the 8th International Conference on Logic for Programming, Artificial Intelli-*

gence and Reasoning (LPAR'01), LNAI No. 2250, pages 100–109. Springer Verlag, 2001.

[2] Natasha Alechina and Brian Logan. Ascribing beliefs to resource bounded agents. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, volume 2, pages 881–888, Bologna, July 2002. ACM Press.

[3] Daniel C. Dennett. *The Intentional Stance*. MIT Press, 1987.

[4] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 480–490, 1985.

[5] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Mass., 1995.

[6] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.

[7] Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. Algorithmic knowledge. In Ronald Fagin, editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fifth Conference (TARK 1994)*, pages 255–266. Morgan Kaufmann, San Francisco, 1994.

[8] J. Hintikka. *Knowledge and belief*. Cornell University Press, Ithaca, NY, 1962.

[9] G. Lakemeyer. Steps towards a first-order logic of explicit and implict belief. In J. Y. Halpern, editor, *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the 1986 Conference*, pages 325–340, San Francisco, Calif., 1986. Morgan Kaufmann.

[10] H. J. Levesque. A logic of implicit and explicit belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence, AAAI-84*, pages 198–202. AAAI, 1984.

[11] D. McDermott. Artificial intelligence meets natural stupidity. *ACM Sigart Newsletter*, 57:4–9, 1976.

[12] Robert C. Moore. *Logic and Representation*. Number 39 in CSLI Lecture Notes. CSLI Publications, 1995.

[13] R. Parikh. Knowledge and the problem of logical omniscience. In *Methodologies for Intelligent Systems, Proceedings of the Second International Symposium*, pages 432–439. North-Holland, 1987.

[14] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484, 1991.

[15] S. J. Rosenschein. Formal theories of knowledge in AI and robotics. *New Generation Computing*, 3(4):345–357, 1985.

[16] S. J. Rosenschein and L. P. Kaelbling. A situated view of representation and control. *Artificial Intelligence*, 73:149–173, 1995.

[17] N. Seel. The 'logical omniscience' of reactive systems. In *Proceedings of the Eighth Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'91)*, pages 62–71, Leeds, England, 1991.

[18] M. P. Singh. Know-how. In Michael Wooldridge and Anand Rao, editors, *Foundations of Rational Agency*, pages 81–104. Kluwer Academic, Dordrecht, 1999.

[19] W. van der Hoek, B. van Linder, and J-J. Ch. Meyer. An integrated modal approach to rational agents. In Michael Wooldridge and Anand Rao, editors, *Foundations of Rational Agency*, pages 133–168. Kluwer Academic, Dordrecht, 1999.

[20] Michael Wooldridge. Computationally grounded theories of agency. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*, pages 13–20. IEEE Press, 2000.

[21] Michael Wooldridge and Alessio Lomuscio. A computationally grounded logic of visibility, perception, and knowledge. *Logic Journal of the IGPL*, 9(2):273–288, 2001.