A Computationally Grounded Logic of Visibility, Perception, and Knowledge

MICHAEL WOOLDRIDGE, *Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, United Kingdom. E-mail: M.J.Wooldridge@csc.liv.ac.uk*

ALESSIO LOMUSCIO, *Department of Computing, Imperial College of Science, Technology and Medicine, London SW7 2BZ, United Kingdom. E-mail: A.Lomuscio@doc.ic.ac.uk*

Abstract

$\mathcal{VSK}$ logic is a family of multi-modal logics for reasoning about the information properties of computational agents situated in some environment. Using $\mathcal{VSK}$ logic, we can represent what is *objectively true* of the environment, the information that is *visible*, or *knowable* about the environment, information the agent *perceives* of the environment, and finally, information the agent actually *knows* about the environment. The semantics of $\mathcal{VSK}$ logic are given in terms of a general, automata-like model of agents. In this paper, we prove completeness for an axiomatisation of $\mathcal{VSK}$ logic, and present correspondence results for a number of $\mathcal{VSK}$ interaction axioms in terms of the architectural properties of the agent that they represent. The completeness proof is novel in that we are able to prove completeness *with respect to the automata-like semantics.* In this sense, $\mathcal{VSK}$ logic is said to be computationally grounded. We give an example to illustrate the formalism, and present conclusions and issues for further work.

1 Introduction

When designing an agent to carry out a task in some environment, it is frequently necessary to reason about the *information properties* of the agent and its environment. For example, many tasks depend on an agent being able to *access* certain information in the environment. If this information is not accessible, then we will not be able to implement an agent to carry out the desired task. Similarly, knowing that a particular piece of information is essential for some task gives us a functional requirement for any agent that will carry out the task: the agent's sensors must be capable of *perceiving* this information. Finally, many applications demand the ability to store and reason about information from the environment.

In this paper, we present a logic that allows us to capture such information properties. *$\mathcal{VSK}$ logic* allows us to represent what is *objectively true* of an environment, what is *visible*, or *knowable* about the environment, what an agent *perceives* of the environment, and finally, what the agent actually *knows* about the environment. Syntactically, $\mathcal{VSK}$ logic is a propositional multi-modal logic, containing modalities "$\mathcal{V}$", "$\mathcal{S}$", and "$\mathcal{K}$", where $\mathcal{V}\varphi$ means that the information φ is accessible in the current environment state; $\mathcal{S}\varphi$ means that the agent perceives φ; and $\mathcal{K}\varphi$ means that the

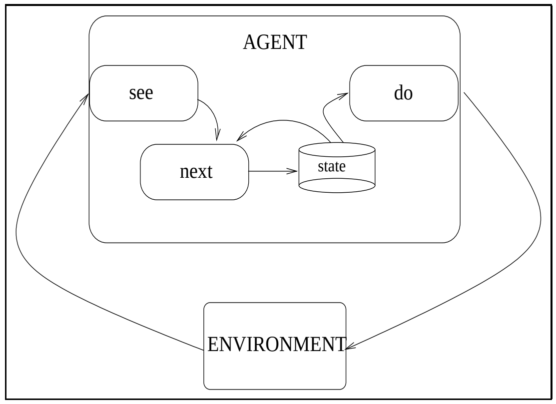

FIG. 1. An overview of the framework.

agent knows φ.

A key feature of the modal system $\mathcal{VSK}$ is that it is computationally grounded [26]. By this we mean that its semantics is not given by traditional possible worlds, but by a formalism that is more closely connected to the models of computation studied in intelligent systems [26]. As we show later in the paper, this does not hinder the possibility of studying formal properties of correspondence [2]; indeed, we are able to show that some axioms formally correspond to intuitive structural properties of the agent/environment computational coupling. Consider, for example, the $\mathcal{VSK}$ formula schema $\mathcal{V}\varphi \Rightarrow \mathcal{S}\varphi$, which says that if the information φ is accessible, then the agent perceives φ. Intuitively, this axiom characterises agents equipped with "perfect" sensors, i.e., sensors that obtain all the information from the environment that is available. In the following, we present results that correspond exactly to this and other intuitions. In addition, we give an axiomatisation of $\mathcal{VSK}$ logic, that we show to be complete with respect to the most general model of agents and environments. In this sense this result is novel. To illustrate the use of $\mathcal{VSK}$ logic, we present a detailed worked example, and we conclude with a discussion of related work and open problems.

2 A Semantic Framework

In this section, we present a semantic model of agents and the environments they occupy. This model plays the role in $\mathcal{VSK}$ logic that *interpreted systems* play in epistemic logic [6, pp103–107] — when we later prove completeness of a $\mathcal{VSK}$ axiomatisation, we prove it with respect to this semantic model. We begin by defining the components modelling the environment; we then define our model of agents; and finally, we combine these to give the notion of a $\mathcal{VSK}$ system. A visual representation of the framework is given in Figure 1.

Following [6], we use the term "environment" to denote all the components of a system external to the agent. Sometimes environments can be represented as just another agent of the system; more often they serve a special purpose, as they can be used to model communication architectures, etc. We model an environment as a

4-tuple containing a set of possible *instantaneous states*, a *visibility function*, which characterises the information content of any given environment state, a *state transformer* function, which characterises the effects that an agent's actions have on the environment, and, finally, an *initial state*.

Definition 2.1 (Environments) *An* environment *is a tuple* $Env = \langle E, vis, \tau_e, e_0 \rangle$, *where:*

- $E = \{e_1, e_2, \ldots\}$ *is a* set of instantaneous local states *for the environment.*
- $vis : E \to 2^E$ *is the* visibility function of the $\mathcal{VSK}$ system. *It is assumed that the function* vis *partitions* E *into mutually disjoint sets and that* $e \in vis(e)$, *for any* $e \in E$. *Elements of the codomain of the function* vis *are called* visibility sets. *We say that* vis *is* transparent *if for any* $e \in E$ *we have that* $vis(e) = \{e\}$.
- $\tau_e : E \times Act \to E$ *is a total* state transformer function *for the environment (cf. [6, p154]), where* Act *is the set of* actions *for the agent (see Definition 2.2). The function* τ_e *is assumed to be an injection.*
- $e_0 \in E$ *is the* initial state *of* Env.

Modelling an environment in terms of a set of states and a state transformer is quite conventional (see, e.g., [6]). One point worthy of note is that we implicitly assume environments evolve *deterministically*: there is no uncertainty about the result of performing an action in some state. The requirement of τ_e being an injection amounts to considering each environment state as unique, i.e., each local state for the environment happens only once in its history. This is equivalent to assuming environments have perfect recall of their history.

The use of the visibility function also requires some explanation. The visibility function defines what is in principle knowable about a $\mathcal{VSK}$ system; the idea is similar to the notion of "partial observability" in POMDPs [12]. Intuitively, not all the information in an environment state is in general accessible to an agent. So $vis(e) = \{e, e', e''\}$ represents the fact that the environment states e, e', e'' are indistinguishable to the agent from e. This is so regardless of the agent's efforts in performing the observation — it represents the maximum amount of information that is in principle available to the agent when observing the state e. The concept of transparency in Definition 2.1 captures "perfect" scenarios, in which all the information in a state is accessible to an agent. Note that visibility functions are *not* intended to capture the everyday notion of visibility as in "object x is visible to the agent".

We adopt a simple, and, we argue, general model of agents, which makes only a minimal commitment to an agent's internal architecture. One important assumption we do make is that agents have an internal state, although we make no assumptions with respect to the actual structure of this state. Agents are assumed to be composed of three functional components: some sensor apparatus, an action selection function, and a next-state function.

Definition 2.2 (Agents) *Given an environment, an* agent *is a tuple* $Ag = \langle L, Act, see, do, \tau_a, \bar{l} \rangle$, *where:*

- $L = \{l_1, l_2, \ldots\}$ *is a set of* instantaneous local states *for the agent.*
- $Act = \{\alpha, \alpha', \ldots\}$ *is a set of* action*s.*

- $see : vis(E) \to Perc$ *is the* perception function*, mapping visibility sets to* percepts. *Elements of the set Perc are denoted as* $\rho, \rho', \ldots$*, and so on. If see is an injection into Perc then we say that see is* perfect*, otherwise we say it is* lossy.
- $do : L \to Act$ *is the* action selection function*, mapping local states to actions.*
- $\tau_a : L \times Perc \to L$ *is the* state transformer function *for the agent. We say that* τ_a *is* complete *if for any global states*[1] $g = (e, \tau_a(l, \rho)), g' = (e', \tau_a(l', \rho'))$ *we have that* $\tau_a(l, \rho) = \tau_a(l', \rho')$ *implies* $\rho = \rho'$*, for every* $l, l' \in L'; e, e' \in E; \rho, \rho' \in Perc$*. We say that* τ_a *is* local *if for any global states* $g = (e, \tau_a(l, \rho))$, $g' = (e', \tau_a(l', \rho))$ *we have that* $\tau_a(l, \rho) = \tau_a(l', \rho)$ *for every* $l, l' \in L; e, e' \in E; \rho \in Perc$.
- $\bar{l} \in L$ *is the* initial state *for the agent.*

Perfect perception functions distinguish between all visibility sets; lossy perception functions are so called because they can map different visibility sets to the same percept, thereby losing information. Note that we have implicitly made the simplifying assumption that the environment evolves synchronously with the agent.

We now require a working definition of the states of a $\mathcal{VSK}$ system, or *global states.*

Definition 2.3 (Global states for a $\mathcal{VSK}$ system) *A set of* global states $G = \{g, g', \ldots\}$ *for a $\mathcal{VSK}$ system is a subset of* $E \times L$.

We do not rule out G being equal to the Cartesian product of E and L; when this happens, the $\mathcal{VSK}$ system is said to be in a *hypercube configuration* and it enjoys some special properties (see [16, 14] for details). We can now define $\mathcal{VSK}$ *systems.*

Definition 2.4 ($\mathcal{VSK}$ systems) *A $\mathcal{VSK}$* system *is a pair* $S = \langle Env, Ag \rangle$*, where Env is an environment, and Ag is an agent. The class of $\mathcal{VSK}$ systems is denoted by* $\mathcal{S}$.

Although the logics we discuss in this paper may be used to refer to *static* properties of knowledge, visibility, and perception, the semantic model naturally allows us to account for the temporal evolution of a $\mathcal{VSK}$ system. The behaviour of an agent situated in an environment can be summarised as follows. The agent starts in state $\bar{l}$, the environment starts in state e_0. At this point the agent "synchronises" with the environment by performing an initial observation through the visibility function vis, and generates a percept $see(vis(e_0))$. The internal state of the agent is then updated, and becomes $l_0 = \tau_a(\bar{l}, see(vis(e_0)))$. The synchronisation phase is now over and the system starts its run from the initial state $g_0 = (e_0, l_0) = (e_0, \tau_a(\bar{l}, see(vis(e_0))))$. An action $\alpha_0 = do(l_0)$ is selected and performed by the agent on the environment, whose state is updated into $e_1 = \tau_e(e_0, \alpha_0)$. The agent enters another cycle, and so on. A *run* of a system is thus a (possibly infinite) sequence of global states defined as follows.

Definition 2.5 (Runs) *A sequence* $(g_0, g_1, g_2, \ldots)$ *over G represents a run of an agent* $Ag = \langle L, Act, see, do, \tau_a, \bar{l} \rangle$ *in an environment* $Env = \langle E, vis, \tau_e, e_0 \rangle$ *if*

- $g_0 = (e_0, l_0), l_0 = \tau_a(\bar{l}, see(vis(e_0)))$.
- *For all u, if* $g_u = (e_u, l_u)$*, then* $g_{u+1} = (e_{u+1}, l_{u+1})$ *is defined by:*

[1] See Definition 2.3.

$$\begin{aligned} e_{u+1} &= \tau_e(e_u, do(l_u)) \quad \textit{and} \\ l_{u+1} &= \tau_a(l_u, see(vis(e_{u+1}))). \end{aligned}$$

- *For any $l \in L$, there exists a g_i such that $g_i = (e_i, l)$. For any $e \in E$, there exists a g_j such that $g_j = (e, l_j)$.*

Note that, since τ_e is an injection, two global states with the same environment component never occur in a run.

Definition 2.6 (Reachable states) *Given a $\mathcal{VSK}$ system $S = \langle Env, Ag \rangle$ we say $G \subseteq E \times L$ is the set of global states* generated *by S whenever $g \in G$ if and only if g occurs in the run of S.*

When $S = \langle Env, Ag \rangle$ is clear from the context we will refer to the set G of global states generated by $S = \langle Env, Ag \rangle$ simply as the set of global states of the $\mathcal{VSK}$ system $S = \langle Env, Ag \rangle$. Note that since both agents and environments are deterministic, a $\mathcal{VSK}$ system has only a single run; in this, we differ from [6].

3 $\mathcal{VSK}$ Logic

We now introduce a language $\mathcal{L}^{\mathcal{VSK}}$, which will enable us to represent the information properties of $\mathcal{VSK}$ systems. In particular, it will allow us to represent first what is true of the $\mathcal{VSK}$ system, then what is *visible*, or *knowable* of the system, then what an agent *perceives* of the system, and finally, what it *knows* of the system.

Definition 3.1 (Syntax of $\mathcal{VSK}$ Logic) *Given a set P of propositional atoms, the language $\mathcal{L}^{\mathcal{VSK}}$ of $\mathcal{VSK}$ logic is defined by the following BNF grammar:*

$$\langle \textit{wff} \rangle ::= \textbf{true} \mid \textit{any element of } P \mid \neg\langle \textit{wff} \rangle \mid \langle \textit{wff} \rangle \wedge \langle \textit{wff} \rangle \mid \mathcal{V}\langle \textit{wff} \rangle \mid \mathcal{S}\langle \textit{wff} \rangle \mid \mathcal{K}\langle \textit{wff} \rangle.$$

The modal operator "$\mathcal{V}$" allows us to represent the information that is instantaneously visible or knowable about the state of the system. Thus, suppose the formula $\mathcal{V}\varphi$ is true in some state $g \in G$. The intended interpretation of this formula is that the property φ is *knowable* of the environment when it is in state g — not only is φ true of the environment, but any agent equipped with suitable sensor apparatus would be able to perceive the information φ. To put it another way, $\mathcal{V}\varphi$ means that an impartial external observer would say that in its current state, the environment carried the information φ. If $\neg\mathcal{V}\varphi$ were true in some state, then *no* agent, no matter how good its sensor apparatus was, would be able to perceive φ.

The fact that something is visible in a $\mathcal{VSK}$ system does not mean that an agent actually sees it. What an agent *does* see is determined by its sensors. The modal operator "$\mathcal{S}$" will be used to represent the information that an agent "sees". The idea is as follows. Suppose an agent's sensory apparatus (represented by the *see* function in our semantic model above) was a video camera, and so the percepts being received by the agent take the form of a video feed. Then $\mathcal{S}\varphi$ means that an impartial observer would say that the video feed currently being supplied by the video camera carried the information φ — in other words, φ is true all situations where the agent received the same video feed.

Finally, $\mathcal{VSK}$ logic allows us to represent an agent's *knowledge*. We represent knowledge by means of a third modal operator, "$\mathcal{K}$". In line with the tradition that started with Hintikka [9], we write $\mathcal{K}\varphi$ to represent the fact that the agent has knowledge of the formula represented by φ. Our model of knowledge is that popularised by Halpern and colleagues [6]: an agent is said to know φ when in local state l, if φ is guaranteed to be true whenever the agent is in state l. As with the $\mathcal{V}$ and $\mathcal{S}$ modalities, knowledge is an *external* notion — an agent is said to know φ if an impartial, omniscient observer would say that the agent's state carried the information φ.

We now proceed to interpret our formal language. While it is entirely possible to do so directly with respect to $\mathcal{VSK}$ systems, we will find it beneficial to use Kripke semantics [13] in order to prove completeness of an axiomatisation. In particular, we will use Kripke frames defined by three relations on their support set.

Definition 3.2 (Kripke frames and models)
A frame *F is a tuple $F = \langle W, R_{\mathcal{V}}, R_{\mathcal{S}}, R_{\mathcal{K}}\rangle$, where W is a non-empty set (whose elements are called* worlds*), and $R_{\mathcal{V}}, R_{\mathcal{S}}, R_{\mathcal{K}} \subseteq W \times W$ are binary relations on W. If all relations are equivalence relations, the frame is an* equivalence frame *and we write $\sim_v, \sim_s, \sim_k$ for $R_{\mathcal{V}}, R_{\mathcal{S}}, R_{\mathcal{K}}$.*

We can define a mapping from the class of $\mathcal{VSK}$ systems to the class of Kripke frames and we can make use of these images to interpret our formal language.

Definition 3.3 (Generated Kripke structures)
Given a $\mathcal{VSK}$ system $S = \langle Env, Ag\rangle$, the Kripke frame *$F_S = \langle W, \sim_v, \sim_s, \sim_k\rangle$* generated by *$S$ is defined as follows:*

- *$W = G$, where G is the set of global states reachable by the system S,*
- *$\sim_v$ is defined by: $(e, l) \sim_v (e', l')$ if $e' \in vis(e)$,*
- *$\sim_s$ is defined by: $(e, l) \sim_s (e', l')$ if $see(vis(e)) = see(vis(e'))$,*
- *$\sim_k$ is defined by: $(e, l) \sim_k (e', l')$ if $l = l'$.*

The class of frames generated by the class of $\mathcal{VSK}$ system $\mathcal{S}$ will be denoted by $\mathcal{F}_{\mathcal{S}}$; similarly F_S will denote the frame generated by the system S. As might be expected, the generated frames are equivalence frames.

Lemma 3.4 *Given any $\mathcal{VSK}$ system $S \in \mathcal{S}$, the frame F_S generated by S is an equivalence frame.*

With Definition 3.3 we have effectively built a bridge between $\mathcal{VSK}$ systems and Kripke frames. In what follows, we assume the standard definitions of satisfaction and validity for Kripke frames and Kripke models defined by three relations on the support set — we refer the reader to [11, 8] for a detailed exposition of the subject. Following [6] and [14], we define the concepts of truth and validity on Kripke models that are *generated* by $\mathcal{VSK}$ systems.

Definition 3.5 (Satisfaction on $\mathcal{VSK}$ systems) *Given an interpretation $\pi : W \to 2^P$, we say that a formula $\varphi \in \mathcal{L}^{\mathcal{VSK}}$ is satisfied at a point $g \in G$ on a $\mathcal{VSK}$ system S if the model $M_S = \langle F_S, \pi\rangle$ built on the generated frame F_S by use of π is such that $M_S \models_g \varphi$. The propositional connectives are assumed to be interpreted as usual and the modal operators $\mathcal{V}, \mathcal{S}, \mathcal{K}$ are assumed to be interpreted in the standard way (see for example [11]) by means of the equivalence relations $\sim_v, \sim_s$, and $\sim_k$ respectively.*

We are especially interested in the properties of a $\mathcal{VSK}$ system as a whole. The notion of validity is appropriate for this analysis.

Definition 3.6 (Validity on $\mathcal{VSK}$ systems) *A formula $\varphi \in \mathcal{L}^{\mathcal{VSK}}$ is valid on a $\mathcal{VSK}$ system S if $F_S \models \varphi$. A formula $\varphi \in \mathcal{L}^{\mathcal{VSK}}$ is valid on a class $\mathcal{T}$ of $\mathcal{VSK}$ systems if for any system $S \in \mathcal{T}$, we have that $F_S \models \varphi$.*

4 Axiomatising $\mathcal{VSK}$ Systems

In this section we study various $\mathcal{VSK}$ systems from an axiomatic perspective. This analysis will let us explore in more detail the properties of visibility, knowledge, and perception of $\mathcal{VSK}$ systems. We begin by presenting correspondence results; we then report completeness of an axiomatisation with respect to the most general class of $\mathcal{VSK}$ systems. We do not report proofs for the correspondence results which are presented in [29].

Let us first note that the class $\mathcal{F}_S$ of frames generated by $\mathcal{VSK}$ systems is a proper subclass of equivalence frames. Indeed, the following holds.

Lemma 4.1 *For any frame $F \in \mathcal{F}_S$, we have $\sim_v \subseteq \sim_s$.*

Lemma 4.2 *$F_S \models \mathcal{S}p \Rightarrow \mathcal{V}p$ if and only if $\sim_v \subseteq \sim_s$.*

In view of these lemmas, any $\mathcal{VSK}$ system validates the formula below.

Corollary 4.3 *Given any S we have $S \models \mathcal{S}p \Rightarrow \mathcal{V}p$.*

Corollary 4.3 is in line with our intuitions about visibility and perception: it says that an agent cannot see something that it is not visible.

We now proceed to give basic correspondence results (see [3] for a detailed exposition of the subject) for axioms relating visibility, perception, knowledge with respect to the architectural classes of $\mathcal{VSK}$ systems described in Section 2. Note that our correspondence results are not simply given with respect to the Kripke frames but to architectural features of $\mathcal{VSK}$ systems.

Lemma 4.4 *1. $S \models p \Rightarrow \mathcal{V}p$ if and only if the system S is transparent.*

2. $S \models p \Rightarrow \mathcal{S}p$ if and only if the system S is transparent and the perception function of the agent Ag in S is perfect.

Lemma 4.4 makes precise the intuition given in the semantics of $\mathcal{VSK}$ systems about transparency and perfect perception. In particular, in order for the agent to be able to perceive everything that is true, it is not enough for it to have a perfect perception function: it also needs to inhabit a system with a transparent visibility function.

We now investigate interaction axioms between visibility, perception, and knowledge. First, recall from Corollary 4.3 that on any generated frame the implication between perception and visibility is valid. Here we turn to the converse direction: if a fact is visible, then it is seen by the agent — in other words, the agent sees everything visible. Intuitively, this axiom characterises agents with "perfect" sensory apparatus, i.e., a *see* function that *never loses information.* Indeed, as the next lemma shows, this axiom corresponds formally to the perception function of the agent being perfect (as defined in Definition 2.2).

Lemma 4.5 *$S \models \mathcal{V}p \Rightarrow \mathcal{S}p$ if and only if the perception function see of the agent Ag in S in perfect.*

Given Corollary 4.3, we can strengthen the above as follows.

Corollary 4.6 *$S \models \mathcal{V}p \Leftrightarrow \mathcal{S}p$ if and only if the perception function see of the agent Ag in S in perfect.*

Suppose we have an agent which assumes that if it cannot see φ, then φ must be false. Such an agent is employing a kind of strict closed world assumption. We formally analyse the contrapositive of it.

Lemma 4.7 *$S \models \neg\mathcal{S}\neg p \Rightarrow \mathcal{V}p$ if and only if system S is transparent and the visibility function is perfect.*

We now turn to the relationship between what an agent perceives and what it knows. Recall from Definition 2.2 that complete transformer functions characterise agents that never lose information when they update their internal state. The following holds.

Lemma 4.8 *$S \models \mathcal{S}p \Rightarrow \mathcal{K}p$ if and only if the state transformer function τ_a is complete.*

Suppose that an agent's internal state at any moment is determined *solely* by the percept it receives at that moment — the agent chooses its next state by ignoring its current local state, and only taking into account the percept that it is currently receiving. This is the *locality* property of the state transformer function τ_a as described in Definition 2.2. For such agents, knowledge is determined solely by the current state of the environment. Indeed, we have the following.

Lemma 4.9 *$S \models \mathcal{K}p \Rightarrow \mathcal{S}p$ if and only if the state transformer function τ_a of system S is local.*

So far we have identified certain classes of $\mathcal{VSK}$ systems. In particular we were able to report that some architectural features of particular $\mathcal{VSK}$ systems are reflected in the validity of some axioms expressing implications between visibility, perception, and knowledge. We now turn our attention to the issue of completeness.

Many different $\mathcal{VSK}$ systems are worth exploring. As discussed above, the environment can be transparent or not, the agent's perception function can be perfect or otherwise, the agent's next state function can be complete, local or neither of the two, and so on. While we reported correspondence results, these are in general not enough to provide completeness, and each semantic class needs its own appropriate analysis. In this article we focus on a basic $\mathcal{VSK}$ logic: we prove that this logic axiomatises the most general class of $\mathcal{VSK}$ systems.

Definition 4.10 *The logic L_{VSK} is the set of formulas generated by the following axiomatisation.*

Taut	$\vdash_{\mathrm{L}_{VSK}} p$, *where p is any propositional tautology*
$K_\mathcal{K}$	$\vdash_{\mathrm{L}_{VSK}} \mathcal{K}(p \Rightarrow q) \Rightarrow (\mathcal{K}p \Rightarrow \mathcal{K}q)$
$T_\mathcal{K}$	$\vdash_{\mathrm{L}_{VSK}} \mathcal{K}p \Rightarrow p$

$4_\mathcal{K}$	$\vdash_{L_{VSK}} \mathcal{K}p \Rightarrow \mathcal{KK}p$
$5_\mathcal{K}$	$\vdash_{L_{VSK}} \neg\mathcal{K}p \Rightarrow \mathcal{K}\neg\mathcal{K}\neg p$
$K_\mathcal{V}$	$\vdash_{L_{VSK}} \mathcal{V}(p \Rightarrow q) \Rightarrow (\mathcal{V}p \Rightarrow \mathcal{V}q)$
$T_\mathcal{V}$	$\vdash_{L_{VSK}} \mathcal{V}p \Rightarrow p$
$4_\mathcal{V}$	$\vdash_{L_{VSK}} \mathcal{V}p \Rightarrow \mathcal{VV}p$
$5_\mathcal{V}$	$\vdash_{L_{VSK}} \neg\mathcal{V}p \Rightarrow \mathcal{V}\neg\mathcal{V}\neg p$
$K_\mathcal{S}$	$\vdash_{L_{VSK}} \mathcal{S}(p \Rightarrow q) \Rightarrow (\mathcal{S}p \Rightarrow \mathcal{S}q)$
$T_\mathcal{S}$	$\vdash_{L_{VSK}} \mathcal{S}p \Rightarrow p$
$4_\mathcal{S}$	$\vdash_{L_{VSK}} \mathcal{S}p \Rightarrow \mathcal{SS}p$
$5_\mathcal{S}$	$\vdash_{L_{VSK}} \neg\mathcal{S}p \Rightarrow \mathcal{S}\neg\mathcal{S}\neg p$
$Int_{\mathcal{S}-\mathcal{V}}$	$\vdash_{L_{VSK}} \mathcal{S}p \Rightarrow \mathcal{V}p$
US	*If* $\vdash_{L_{VSK}} \varphi$, *then* $\vdash_{L_{VSK}} \varphi[\psi_1/p_1, \ldots, \psi_n/p_n]$
MP	*If* $\vdash_{L_{VSK}} \varphi$ *and* $\vdash_{L_{VSK}} \varphi \Rightarrow \psi$, *then* $\vdash_{L_{VSK}} \psi$
$Nec_\mathcal{K}$	*If* $\vdash_{L_{VSK}} \varphi$, *then* $\vdash_{L_{VSK}} \mathcal{K}\varphi$
$Nec_\mathcal{V}$	*If* $\vdash_{L_{VSK}} \varphi$, *then* $\vdash_{L_{VSK}} \mathcal{V}\varphi$
$Nec_\mathcal{S}$	*If* $\vdash_{L_{VSK}} \varphi$, *then* $\vdash_{L_{VSK}} \mathcal{S}\varphi$

It is immediately apparent that each of the $\mathcal{VSK}$ modalities enjoy the properties of an S5 modal logic: they each validate analogues of the modal logic axioms KT45 [11, 8]. The appropriateness of S5 as a logic of (idealised) knowledge has been discussed at length in the literature, and is now widely accepted [6, pp30–36]; for this reason, we will not motivate the S5 logic of knowledge. However, the appropriateness of S5 for the $\mathcal{V}$ and $\mathcal{S}$ modalities requires some justification.

Consider the $\mathcal{V}$ modality first. Recall the intended interpretation of a formula $\mathcal{V}\varphi$: that $\mathcal{V}\varphi$ is true in some state if an impartial observer would say that this state carried the information φ. Taking the axioms KT45 in turn, $K_\mathcal{V}$ seems unproblematic: if the information $p \Rightarrow q$ and p is carried by a state, then q must also be carried by that state. Axiom $T_\mathcal{V}$ simply says that if information p is carried by a state, then p must be true. This is a desirable property, since it would seem unreasonable to say that a state really carried some information if that information were false. Axiom $4_\mathcal{V}$ says that if we can conclude that a state carries information p, then we also have some additional (although arguably not terribly helpful) information: that it carries the information that it carries the information p. Since we have axiom $T_\mathcal{V}$, it follows that $\mathcal{V}p \Leftrightarrow \mathcal{VV}p$ will be an axiom: we can remove repeated occurrences of the $\mathcal{V}$ modality without affecting the truth of a formula. Finally, axiom $5_\mathcal{V}$ says that if we can conclude that a state does not carry the information p, then we can conclude that the state carries the information that it does not carry the information p. Axioms $4_\mathcal{V}$ and $5_\mathcal{V}$ thus extend our information about a state from understanding the limits to the information carried by that state.

Turning to the $\mathcal{S}$ modality, we should first emphasise that $\mathcal{S}$ is *not* intended to form a logic of perception in the sense of, for example, Hintikka's [10, pp151–183]. Rather, $\mathcal{S}$ captures an *objective* notion of perception, (what an omniscient impartial observer would say you are seeing), rather than a *subjective* view of perception (what you believe you are seeing). Thus $\mathcal{S}\varphi$ means that if the agent is receiving some percept ρ, then whenever it receives percept ρ, formula φ is guaranteed to be true. In this sense, the percept the agent receives is carrying the information φ. We argue that under this interpretation, the S5 axioms capture reasonable properties of the $\mathcal{S}$ modality. The most controversial of these axioms for $\mathcal{S}$ is $T_\mathcal{S}$, and it is therefore worth examining this axiom in more detail. It says that if an agent "sees" p, then p must be true. If we were attempting to capture the everyday sense of *human* perception, then this axiom would not be acceptable — there are many obvious reasons why, if you perceive p, you could be wrong. However, under our interpretation, we say that $\mathcal{S}p$ means that in every state where you receive the same percept that you are currently receiving, p is true — in particular, p must be true in the current state. We can argue similarly for axioms $4_{\mathcal{S}}$ and $5_{\mathcal{S}}$.

We can prove that Definition 4.10 represents a sound and complete axiomatisation for the most general class of $\mathcal{VSK}$ systems.

Theorem 4.11 *The logic* L_{VSK} *is sound and complete with respect to the class of* $\mathcal{VSK}$ *systems* $\mathcal{S}$.

PROOF. (Outline) It is straightforward to show that L_{VSK} is sound with respect to the class of systems $\mathcal{S}$. In order to prove completeness, it suffices to show that $\mathrm{L}_{VSK} \not\vdash \varphi$ implies $\mathcal{S} \not\models \varphi$, for any $\varphi \in \mathcal{L}^{\mathcal{VSK}}$. By carrying out a routine proof via the canonical model method (cf., e.g., [20]) one can show that the logic L_{VSK} is complete with respect to the class $\mathcal{G}$ of equivalence frames $F = \langle U, \sim_v, \sim_s, \sim_k \rangle$, where $\sim_v \subseteq \sim_s$. But it can be proven that given any frame $G = \langle W, \sim_v, \sim_s, \sim_k \rangle \in \mathcal{G}$, one can define a system $S \in \mathcal{S}$ such that its generated frame F_S is the domain of a p-morphism onto G. The construction of S is somehow cumbersome and not reported here (see [29] for details).

Suppose then $\mathrm{L}_{VSK} \not\vdash \varphi$, then by the completeness result above we have $G \not\models \varphi$ for some $G \in \mathcal{G}$; but then by constructing S as suggested above, we can prove that, because of considerations on the transfer of validity to p-morphic images (e.g., see [8, page 11] for details on the mono-modal case), $F_S \not\models \varphi$. So $S \not\models \varphi$, hence $\mathcal{S} \not\models \varphi$, where $\mathcal{S}$ is the class of $\mathcal{VSK}$ systems. ■

5 A Case Study

To illustrate the formalism, we present a simple case study, adapted from Russell and Norvig's vacuum world [24, p58]. A robot agent occupies an environment with two rooms, room 1 and room 2. The rooms are connected by a single door, which may be open or closed. Initially, the robot is in room 1. There may be dirt on the floor in either or both of these rooms. The robot can detect whether the door is open or closed, and whether it is in the same room as some dirt, but that is all. It has a vacuum cleaner, which, if operated, will suck dirt from underneath it. It is also capable of opening the door and moving from one room to the other. When the door is closed, it is *impossible* to tell whether there is dirt in the other room. However, when the door is open, an agent with sufficiently powerful sensory equipment *could* in principle detect dirt in the other room. Our agent is not capable of this. To characterise the properties of the environment, we use five primitive propositions: d_i (where $i = 1, 2$) indicates that dirt is in room i; ag_i (where $i = 1, 2$) indicates that the agent is in room i; and *door* indicates whether the door is open. The possible states of the environment are $E = \{e_0, \ldots, e_15\}$, and are summarised in Figure 2. Note that we encode in the environment the physical position of the agent. We discuss how to code its internal state later on.

We can represent the environment's visibility function, vis, as follows:

State	d_1?	d_2?	ag_1?	ag_2?	$door$?
e_0	N	N	Y	N	N
e_1	N	Y	Y	N	N
e_2	N	N	Y	N	Y
e_3	N	Y	Y	N	Y
e_4	Y	N	Y	N	N
e_5	Y	Y	Y	N	N
e_6	Y	N	Y	N	Y
e_7	Y	Y	Y	N	Y
e_8	N	N	N	Y	N
e_9	N	Y	N	Y	N
e_{10}	N	N	N	Y	Y
e_{11}	N	Y	N	Y	Y
e_{12}	Y	N	N	Y	N
e_{13}	Y	Y	N	Y	N
e_{14}	Y	N	N	Y	Y
e_{15}	Y	Y	N	Y	Y

TABLE 1. Possible states of the Vacuum World.

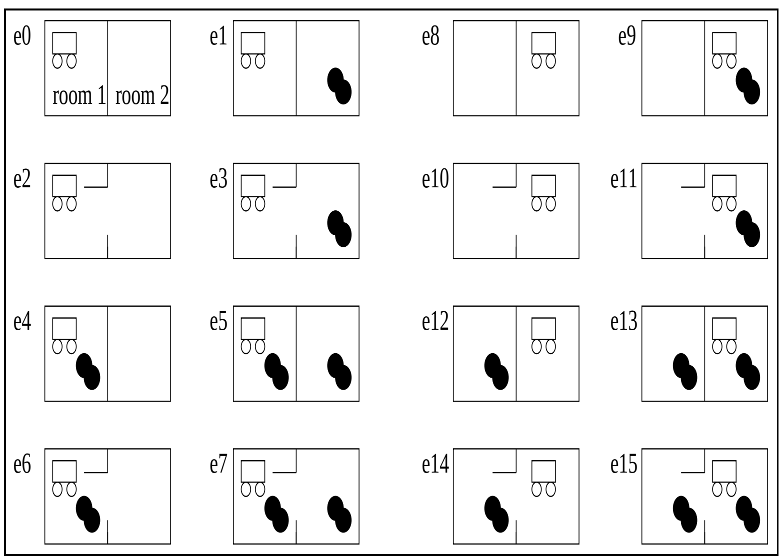

FIG. 2. The vacuum world.

$$vis(e_i) = \begin{cases} \{e_0, e_1\} & \text{if } e_i = e_0 \text{ or } e_i = e_1 \\ \{e_4, e_5\} & \text{if } e_i = e_4 \text{ or } e_i = e_5 \\ \{e_8, e_{12}\} & \text{if } e_i = e_8 \text{ or } e_i = e_{12} \\ \{e_9, e_{13}\} & \text{if } e_i = e_9 \text{ or } e_i = e_{13} \\ \{e_i\} & \text{otherwise.} \end{cases}$$

The *see* function for the agent is defined as:

$$see(X) = \begin{cases} \rho_0 & \text{if } X = \{e_0, e_1\} \text{ or } X = \{e_8, e_{12}\} \\ \rho_1 & \text{if } X = \{e_4, e_5\} \text{ or } X = \{e_9, e_{13}\} \\ \rho_2 & \text{if } X \in \{\{e_2\}, \{e_3\}, \{e_{10}\}, \{e_{14}\}\} \\ \rho_3 & \text{if } X \in \{\{e_6\}, \{e_7\}, \{e_{11}\}, \{e_{15}\}\} \end{cases}$$

Notice that *see* is lossy. We leave the reader to see that:

- Percept ρ_0 carries the information that the door is closed and either the agent is in room 1 and there is no dirt in room 1, or the agent is in room 2 and there is no dirt in room 2, i.e.,

$$\neg door \wedge ((ag_1 \wedge \neg d_1) \vee (ag_2 \wedge \neg d_2)) \tag{5.1}$$

  Notice that the percept does not indicate which room the agent is in.
- Percept ρ_1 carries the information that the door is closed and that either the agent is in in room 1 and there is dirt in room 1, or the agent is in room 2 and there is dirt in room 2, i.e.,

$$\neg door \wedge ((ag_1 \wedge d_1) \vee (ag_2 \wedge d_2)) \tag{5.2}$$

- Percept ρ_2 carries the information that the door is open and that either the agent is in in room 1 and there is no dirt in room 1, or the agent is in room 2 and there is no dirt in room 2, i.e.,

$$door \wedge ((ag_1 \wedge \neg d_1) \vee (ag_2 \wedge \neg d_2)) \tag{5.3}$$

- Percept ρ_3 carries the information that the door is open and that either the agent is in in room 1 and there is dirt in room 1, or the agent is in room 2 and there is dirt in room 2, i.e.,

$$door \wedge ((ag_1 \wedge d_1) \vee (ag_2 \wedge d_2)) \tag{5.4}$$

We now describe the evolution of the agent. First we need to define its local states. For the sake of this example, all we need to code is the representation of the environment in the internal states of the agent. To this end we consider a set $L = \{\bar{l}, l_0, l_1, l_2, l_3\}$, where $\bar{l}$ is the initial state, l_0 the state coding the door being closed, and no dirt being present; l_1 the state corresponding to the door being closed, and dirt being present; l_2 the state referring to the door being open, and no dirt being present; l_3 the state in which the door is supposed open, and some dirt is present. The evolution function τ_a function is defined as:

$$\tau_a(l_i, \rho_j) = \begin{cases} l_0 & \text{if } \rho_j = \rho_0 \\ l_1 & \text{if } \rho_j = \rho_1 \\ l_2 & \text{if } \rho_j = \rho_2 \\ l_3 & \text{if } \rho_j = \rho_3 \end{cases}$$

Notice that τ_a is local. The *do* function is defined:

$$\underbrace{\langle e_5, l_2\rangle}_{g_0} \stackrel{suck}{\rightarrow} \underbrace{\langle e_1, l_1\rangle}_{g_1} \stackrel{open}{\rightarrow} \underbrace{\langle e_3, l_3\rangle}_{g_2} \stackrel{move}{\rightarrow} \underbrace{\langle e_{11}, l_4\rangle}_{g_3} \stackrel{suck}{\rightarrow} \underbrace{\langle e_{10}, l_3\rangle}_{g_4} \stackrel{move}{\rightarrow} \underbrace{\langle e_2, l_3\rangle}_{g_5} \stackrel{move}{\rightarrow} \cdots$$

FIG. 3. The history of the vacuum world.

$$do(l) = \begin{cases} null & \text{if } l = \bar{l} \\ open & \text{if } l = l_0 \\ move & \text{if } l = l_2 \\ suck & \text{if } l = l_1 \text{ or } l = l_3 \end{cases}$$

Define now the $\mathcal{VSK}$ system $S = \langle Env, Ag\rangle$, with environment and agent as defined above. Suppose that initially, there is dirt in both rooms, but the door is closed; i.e., that the environment begins in state e_5. Agent and environment, then synchronise producing the initial state $g_0 = (e_5, l_2)$. Then we can deterministically compute the run traced out by the agent when placed in the environment: this is pictured in Figure 3.

Define now the generated model $M_S = \langle F_s, \pi\rangle$, where F_S is the frame generated from system S according to Definition 3.3, and π is an interpretation for the atoms that complies to that described in Figure 1. It is now easy to verify that (for brevity, and by abusing the syntax we use the labels of the formulas above, rather than copying them):

Observation 5.1 $M_S \models_{g_0} d_1 \wedge d_2 \wedge \mathcal{V}d_1 \wedge \neg\mathcal{V}d_2 \wedge \mathcal{S}(5.2) \wedge \mathcal{K}(5.2)$

In state g_1, there is no longer any dirt in room 1. We can see that this situation corresponds to the following.

Observation 5.2 $M_S \models_{g_1} (\neg d_1) \wedge d_2 \wedge \mathcal{V}\neg d_1 \wedge \neg\mathcal{V}d_2 \wedge \mathcal{S}(5.1) \wedge \mathcal{K}(5.1)$

The agent then performs the *open* action. The fact that there is dirt in room 2 then becomes visible, even though the agent does not see it.

Observation 5.3 $M_S \models_{g_2} (\neg d_1) \wedge d_2 \wedge \mathcal{V}(d_2 \wedge \neg d_1) \wedge \mathcal{S}(5.3) \wedge \mathcal{K}(5.3)$

The agent then executes a *move* action, moving to room 2.

Observation 5.4 $M_S \models_{g_3} (\neg d_1) \wedge d_2 \wedge (\mathcal{V}d_2 \wedge \neg d_1) \wedge \mathcal{S}(5.4) \wedge \mathcal{K}(5.4)$

After the agent performs a "*suck*" action the environment is transformed to e_{10}. We can characterise the salient features of g_4 as follows.

Observation 5.5 $M_S \models_{g_4} \neg(d_1 \vee d_2) \wedge \mathcal{V}\neg(d_1 \vee d_2) \wedge \mathcal{S}(5.3) \wedge \mathcal{K}(5.3)$

We leave it to the reader to characterise state g_5 similarly.

6 Related Work

Since the mid 1980s, Halpern and colleagues have used modal epistemic logic for reasoning about multi-agent systems [6]. In this work, they demonstrated how *interpreted systems* could be used as models for such logics. Interpreted systems are very close to our agent-environment systems: the key differences are that they only record the *state* of agents within a system, and hence do not represent the percepts received by an agent or distinguish between what is true of an environment and what is visible of that environment. Halpern and colleagues have established a range of significant results relating to such logics, in particular, categorisations of the complexity of various decision problems in epistemic logic, the circumstances under which it is possible for a group of agents to achieve "common knowledge" about some fact, and most recently, the use of such logics for *directly programming* agents [7]. Comparatively little effort has been devoted to characterising "architectural" properties of agents. The only obvious examples are the temporal properties of no learning, perfect recall, and so on [6, pp281–307]. In their "situated automata" paradigm, Kaelbling and Rosenschein directly synthesised agents (in fact, digital circuits) from epistemic specifications of these agents [23]. While this work clearly highlighted the relationship between epistemic theories of agents and their realisation, it did not explicitly investigate axiomatic characterisations of architectural agent properties. Finally, recent work has considered knowledge-theoretic approaches to robotics [4].

Many other formalisms for reasoning about intelligent agents and multi-agent systems have been proposed over the past decade [28]. Following the pioneering work of Moore on the interaction between knowledge and action [19], most of these formalisms have attempted to characterise the "mental state" of agents engaged in various activities. Well-known examples of this work include Cohen-Levesque's theory of intention [5], and the ongoing work of Rao-Georgeff on the belief-desire-intention (BDI) model of agency [21, 27]. The emphasis in this work has been more on axiomatic characterisations of architectural properties; for example, in [22], Rao-Georgeff discuss how various axioms of BDI logic can be seen to intuitively correspond to properties of agent architectures. However, this work is specific to BDI architectures, and in addition, the correspondence is an *intuitive* one: they establish no formal correspondence, in the sense of $\mathcal{VSK}$ logic. In this sense, BDI logics (and most of their close relatives) are not computationally grounded. The notion of computationally grounded theories of agency is discussed in [26].

A number of author have considered the problem of reasoning about actions that may be performed in order to obtain information. Again building on the work of Moore [19], the goal of such work is typically to develop representations of sensing actions that can be used in planning algorithms [1]. An example is [25], in which Scherl and Levesque develop a representation of sensing actions in the situation calculus [17]. These theories focus on giving an account of how the performance of a sensing action changes an agent's knowledge state. Such theories are purely axiomatic in nature — no architectural correspondence is established between axioms and models that they correspond to.

7 Conclusions

In this paper, we have introduced $\mathcal{VSK}$ logic as a formalism for representing and reasoning about the information properties of agents and their environments. Using $\mathcal{VSK}$ logic, we are able to represent what is objectively true of some environment, what is accessible or visible of the environment, what an agent sees of the environment, and finally, what an agent knows. The semantics of $\mathcal{VSK}$ logic were presented with respect to a simple and general model of agents and their environments. We were able to prove correspondence results for a number of possible axioms of $\mathcal{VSK}$ logic with respect to this model of agents and environments, thus demonstrating that certain axioms captured quite intuitive architectural properties of agent/environment systems. Finally, we gave an axiomatisation of $\mathcal{VSK}$ logic, and proved completeness of this logic with respect to the formal model of agents and environments. It is worth stressing that completeness was shown with the *grounded* semantics of Section 2 and that Kripke models are only used as a vehicle to achieve the result.

There are many avenues for future work: temporal extensions and multi-agent extensions are two of the most important. Completeness results for all basic $\mathcal{VSK}$ systems are another area of work. Finally, decidability and complexity results are desirable, perhaps by using the results of [6, pp62–76].

Finally, it is worth noting that other formalisms — such as [18, 15] — have been explored to represent automata-like models of computation on which epistemic languages can be defined in similar way. Although the notion of perception can be represented there, it is not clear that also the notion of visibility can also be coded. A comparison between this and the formalism of this paper is clearly a topic for further investigation.

References

[1] J. F. Allen, J. Hendler, and A. Tate, editors. *Readings in Planning.* Morgan Kaufmann Publishers: San Mateo, CA, 1990.

[2] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, volume 165 of *Synthese Library*, chapter II.4, pages 167–247. D. Reidel Publ. Co., Dordrecht, 1984.

[3] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 167–247. D. Reidel Publ. Co.: Dordrecht, 1984.

[4] R. I. Brafman and Y. Shoham. Knowledge considerations in robotics and distribution of robotic tasks. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 96–102, Montréal, Québec, Canada, 1995.

[5] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[6] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge.* The MIT Press: Cambridge, MA, 1995.

[7] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. Knowledge-based programs. *Distributed Computing*, 10(4):199–225, 1997.

[8] R. Goldblatt. *Logics of Time and Computation (CSLI Lecture Notes Number 7).* Center for the Study of Language and Information, Ventura Hall, Stanford, CA 94305, 1987. (Distributed by Chicago University Press).

[9] J. Hintikka. *Knowledge and Belief.* Cornell University Press: Ithaca, NY, 1962.

[10] J. Hintikka. *Models for Modalities.* Kluwer Academic Publishers: Boston, MA, 1969.

[11] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic.* Routledge: New York, 1996.

[12] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

[13] S. Kripke. Semantical analysis of modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9:67–96, 1963.

[14] A. Lomuscio. *Knowledge Sharing among Ideal Agents.* PhD thesis, School of Computer Science, University of Birmingham, Birmingham, UK, June 1999.

[15] A. Lomuscio, R. van der Meyden, and M. Ryan. Knowledge in multi-agent systems: Initial configurations and broadcast. *ACM Transactions of Computational Logic*, 1(2), October 2000.

[16] A. Lomuscio and M. Ryan. Ideal agents sharing (some!) knowledge. In *Proceedings of the Thirteenth European Conference on Artificial Intelligence (ECAI-98)*, pages 557–561, Brighton, United Kingdom, August 1998.

[17] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*. Edinburgh University Press, 1969.

[18] R. van der Meyden. Knowledge based programs: On the complexity of perfect recall in finite environments. In Y. Shoham, editor, *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 31–50, San Francisco, March17–20 1996. Morgan Kaufmann Publishers.

[19] R. C. Moore. A formal theory of knowledge and action. In J. F. Allen, J. Hendler, and A. Tate, editors, *Readings in Planning*, pages 480–519. Morgan Kaufmann Publishers: San Mateo, CA, 1990.

[20] S Popkorn. *First Steps in Modal Logic.* Cambridge University Press: Cambridge, England, 1994.

[21] A. S. Rao and M. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–344, 1998.

[22] A. S. Rao and M. P. Georgeff. An abstract architecture for rational agents. In C. Rich, W. Swartout, and B. Nebel, editors, *Proceedings of Knowledge Representation and Reasoning (KR&R-92)*, pages 439–449, 1992.

[23] S. J. Rosenschein and L. P. Kaelbling. A situated view of representation and control. In P. E. Agre and S. J. Rosenschein, editors, *Computational Theories of Interaction and Agency*, pages 515–540. The MIT Press: Cambridge, MA, 1996.

[24] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach.* Prentice-Hall, 1995.

[25] R. Scherl and H. Levesque. The frame problem and knowledge-producing actions. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, pages 689–695, Washington DC, 1993.

[26] M. Wooldridge. Computationally grounded theories of agency. In *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*, pages 13–20, Boston, MA, 2000.

[27] M. Wooldridge. *Reasoning about Rational Agents.* The MIT Press: Cambridge, MA, 2000.

[28] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.

[29] M. Wooldridge and A. Lomuscio. A logic of visibility, perception, and knowledge, 2000. In preparation.

Received November 1, 2000