Reasoning about Visibility, Perception, and Knowledge

Michael Wooldridge and Alessio Lomuscio

Department of Electronic Engineering
Queen Mary and Westfield College
University of London
London E1 4NS, United Kingdom
{M.J.Wooldridge, A.R.Lomuscio}@qmw.ac.uk

Abstract. Although many formalisms have been proposed for reasoning about intelligent agents, few of these have been semantically *grounded* in a concrete computational model. This paper presents $\mathcal{VSK}$ logic, a formalism for reasoning about multi-agent systems, in which the semantics are grounded in an general, finite state machine-like model of agency. $\mathcal{VSK}$ logic allows us to represent: what is *objectively true* of the environment; what is *visible*, or *knowable* about the environment; what the agent *perceives* of the environment; and finally, what the agent actually *knows* about the environment. $\mathcal{VSK}$ logic is an extension of modal epistemic logic. The possible relationships between what is true, visible, perceived, and known are discussed and characterised in terms of the architectural properties of agents that they represent. Some conclusions and issues are then discussed.

1 Introduction

Many formalisms have been proposed for reasoning about intelligent agents and multi-agent systems [16]. However, most such formalisms are *ungrounded*, in the sense that while they have a mathematically well-defined semantics, these semantics cannot be given a *computational* interpretation. This throws doubt on the claim that such logics can be useful for reasoning about computational agent systems.

One formalism that does not fall prey to this problem is epistemic logic — the modal logic of knowledge [5]. Epistemic logic is computationally grounded in that it has a natural interpretation in terms of the states of computer processes. Epistemic logic can be seen as a tool with which to represent and reason about what is *objectively true* of a particular environment and the *information* that agents populating this environment have about it.

Although epistemic logic has proved to be a powerful tool with which to reason about agents and multi-agent systems, it is not expressive enough to capture certain key aspects of agents and their environments. First, there is in general a distinction between what is instantaneously *true* of an environment and what is *knowable* or *visible* about it. To pick an extreme example, suppose p represents the fact that the temperature at the north pole of Mars is 200K. Now it may be that as we write, p is true of the physical world — but the laws of physics prevent us from having immediate access to this information. In this example, something is true in the environment, but this information is *inaccessible*. Traditional epistemic logics can represent p itself, and also allow us to represent the fact that the agent does not know p. But there is no way of distinguishing in normal modal logic between information that is both true and accessible, and statements that are true but not accessible. Whether or not a property is accessible in some environment will have a significant effect on the design of agents to operate in that environment.

In a similar way, we can distinguish between information that is accessible in an environment state, and the information an agent actually perceives of that environment state. For example, it may be that a particular fact is knowable about some environment, but that the agent's sensors are not capable of perceiving this fact. Again, the relationship between what is knowable about an environment and what an agent actually perceives of it has an impact on agent design. Finally, we can also distinguish between the information that an agent's sensors carry and the information that the agent actually carries in its state, i.e., its knowledge.

In this paper, we present a formalism called $\mathcal{VSK}$ logic, which allows us capture these distinctions. $\mathcal{VSK}$ logic allows us to represent what is *objectively true* of the environment, information that is *visible*, or *knowable* about the environment, information the agent *perceives* of the environment, and finally, what the agent actually *knows* about the environment. $\mathcal{VSK}$ logic is an extension of modal epistemic logic. The underlying semantic model is closely related to the interpreted systems model, which is widely used to give a semantics to modal epistemic logic [5, pp103–114]. A key contribution of $\mathcal{VSK}$ logic is that possible relationships between what is true, visible, perceived, and known are characterised model theoretically in terms of the *architectural properties* of agents that they correspond to.

The remainder of this paper is structured as follows. First, the formal model that underpins $\mathcal{VSK}$ logic is presented. In the sections that follow, the logic itself is developed, and some systems of $\mathcal{VSK}$ logic are discussed. An example is presented, illustrating the formalism. Finally, related work is presented, along with some conclusions, and some open issues are briefly discussed. We begin, in the following section, by presenting the underlying semantic model.

2 A Formal Model

In this section, we present a simple formal model of agents and the environments they occupy — see Figure 1 (cf. [6, pp307–313]. We start by introducing the basic sets used in our formal model. First, it is assumed that the environment may be in any of a set $E = \{e, e', \ldots\}$ of states, and that the (single) agent occupying this environment may be in any of a set $L = \{l, l', \ldots\}$ of *local* states. Agents are assumed to have a repertoire of possible actions available to them, which transform the state of the environment — we let $Ac = \{\alpha, \alpha', \ldots\}$ be the set of actions. We assume a distinguished member *null* of Ac, representing the "noop" action, which has no effect on the environment.

In order to represent the effect that an agent's actions have on an environment, we introduce a *state transformer* function, $\tau : E \times Ac \to E$ (cf. [5, p154]). Thus $\tau(e, \alpha)$ denotes the environment state that would result from performing action α in environment state e. Note that our environments are *deterministic*: there is no uncertainty about the

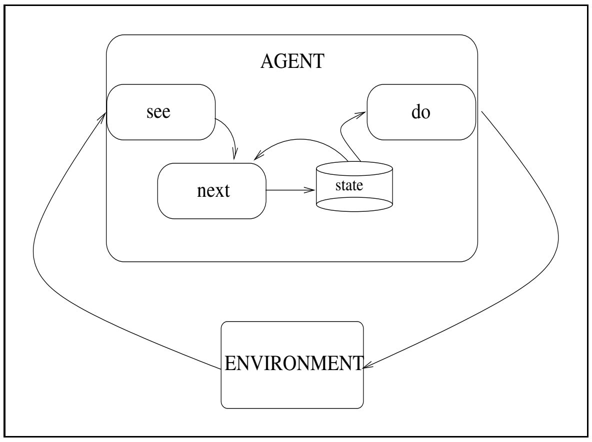

Fig. 1. An overview of the framework.

result of performing an action in some state. Dropping this assumption is not problematic, but it does make the formalism somewhat more convoluted.

In order to represent what is knowable about the environment, we use a *visibility function*, $\nu : E \to (\wp(E) \setminus \emptyset)$. The idea is that if the environment is *actually* in state e, then it is impossible for any agent in the environment to distinguish between e and any member of $\nu(e)$. We require that ν partitions E into mutually disjoint sets of states, and that $e \in \nu(e)$, for all $e \in E$. For example, suppose $\nu(e_2) = \{e_2, e_3, e_4\}$. Then the intuition is that the agent would be unable to distinguish between e_2 and e_3, or between e_2 and e_4. Note that visibility functions are *not* intended to capture the everyday notion of visibility as in "object x is visible to the agent".

We will say ν is *transparent* if $\nu(e) = \{e\}$. Intuitively, if ν is transparent, then it will be possible for an agent observing the environment to distinguish every different environment state.

Formally, an environment *Env* is a 4-tuple $\langle E, \tau, \nu, e_o \rangle$, where E is a set of environment states as above, τ is a state transformer function, ν is a visibility function, and $e_0 \in E$ is the initial state of *Env*.

From Figure 1, we can see that an agent has three functional components, representing its sensors (the function *see*), its next state function (*next*), and its action selection, or decision making function (*do*). Formally, the perception function $see : \wp(E) \to P$ maps sets of environment states to *percepts* — we denote members of P by $\rho, \rho', \ldots$. The agent's next state function $next : L \times P \to L$ maps an internal state and percept to an internal state; and the action-selection function $do : L \to Ac$ simply maps internal states to actions.

The behaviour of an agent can be summarised as follows. The agent starts in some state l_0. It then observes its environment state e_0 through the visibility function $\nu(e_0)$, and generates a percept $see(\nu(e_0))$. The internal state of the agent is then updated to $next(l_0, see(\nu(e_0)))$. The action selected by the agent is then $do(next(l_0, see(\nu(e_0))))$. This action is performed, and the agent enters another cycle.

Together, an environment/agent pair comprise a *system*. The *global* state of a system at any time is a pair containing the state of the agent and the state of the environment. Let $G = E \times L$ be the set of all such global states. We use g (with annotations: $g, g', \ldots$) to stand for members of G. A *run* of a system can be thought of as an infinite sequence:

$$g_0 \xrightarrow{\alpha_0} g_1 \xrightarrow{\alpha_1} g_2 \xrightarrow{\alpha_2} g_3 \xrightarrow{\alpha_3} \cdots \xrightarrow{\alpha_{u-1}} g_u \xrightarrow{\alpha_u} \cdots$$

A sequence $(g_0, g_1, g_2, \ldots)$ over G represents a run of an agent $\langle see, next, do, l_0 \rangle$ in an environment $\langle E, \tau, \nu, e_0 \rangle$ iff:

1. $g_0 = \langle e_0, next(l_0, see(vis(e_0))) \rangle$ and;
2. $\forall u \in I\!N$, if $g_u = \langle e, l \rangle$ and $g_{u+1} = \langle e', l' \rangle$ then

$$\begin{aligned} e' &= \tau(e, do(l)) \quad \text{and} \\ l' &= next(l, e') \end{aligned}$$

Let $G_{Env,Ag} \subseteq G$ denote the set of global states that system Env, Ag could enter during execution.

In order to represent the properties of systems, we assume a set $\Phi = \{p, q, r, \ldots\}$ of primitive propositions. In order to *interpret* these propositions, we use a function $\pi : \Phi \times G_{Ag,Env} \to \{T, F\}$. Thus $\pi(p, g)$ indicates whether proposition $p \in \Phi$ is true (T) or false (F) in state $g \in G$. Note that members of Φ are assumed to express properties of *environment states only*, and *not* the internal properties of agents. We also require that any two different states differ in the valuation of at least one primitive proposition.

We refer to a triple $\langle Env, Ag, \pi \rangle$ as a *model* — our models play the role of interpreted systems in knowledge theory [5, p110]. We use M (with annotations: $M', M_1, \ldots$) to stand for models.

3 Truth and Visibility

Now that we have the formal preliminaries in place, we can start to consider the relationships that we discussed in section 1. We progressively introduce a logic $\mathcal{L}$, which will enable us to represent first what is true of the environment, then what is visible, or knowable of the environment, then what an agent perceives of the environment, and finally, what it knows of the environment.

We begin by introducing the propositional logic fragment of $\mathcal{L}$, which allows us to represent what is true of the environment. Propositional formulae of $\mathcal{L}$ are built up from Φ using the classical logic connectives "$\wedge$" (and), "$\vee$" (or), "$\neg$" (not), "$\Rightarrow$" (implies), and "$\Leftrightarrow$" (if, and only if), as well as logical constants for truth ("**true**") and falsity ("**false**"). We define the syntax and semantics of the truth constant, disjunction, and negation, and assume the remaining connectives and constants are introduced as abbreviations in the conventional way. Formally, the syntax of the propositional fragment of $\mathcal{L}$ is defined by the following grammar:

$$\langle wff \rangle ::= \textbf{true} \mid \text{any element of } \Phi \mid \neg \langle wff \rangle \mid \langle wff \rangle \vee \langle wff \rangle$$

The semantics are defined via the satisfaction relation "$\models$":

$$
\begin{array}{lll}
\langle M, g \rangle \models \textbf{true} & & \\
\langle M, g \rangle \models p & \text{iff } \pi(p, g) = T & (\text{where } p \in \Phi) \\
\langle M, g \rangle \models \neg\varphi & \text{iff not } \langle M, g \rangle \models \varphi & \\
\langle M, g \rangle \models \varphi \vee \psi & \text{iff } \langle M, g \rangle \models \varphi \text{ or } \langle M, g \rangle \models \psi &
\end{array}
$$

We will assume the conventional definitions of satisfiability, validity, and validity in a model.

We now enrich $\mathcal{L}$ by the addition of a unary modality "$\mathcal{V}$", which will allow us to represent the information that is instantaneously visible or knowable about an environment state. 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; in other words, that an agent equipped with suitable sensory apparatus would be able to perceive the information φ. If $\neg\mathcal{V}\varphi$ were true in some state, then *no* agent, no matter how good its sensory apparatus was, would be able to perceive φ.

Note that our concept of visibility is distinct from the everyday notion of visibility as in "object o is visible to the agent". If we were interested in capturing this notion of visibility we could use a first-order logic predicate along the lines of $\mathit{visible}(x, y, o)$ to represent the fact that when an agent is in position (x, y), object o is visible. The arguments to such visibility statements are *terms*, whereas the arguments to the visibility statement $\mathcal{V}\varphi$ is a *proposition*.

In order to give a semantics to the $\mathcal{V}$ operator, we define a binary *visibility accessibility relation* $\sim_\nu \subseteq G_{Ag,Env} \times G_{Ag,Env}$ as follows: $\langle e, l \rangle \sim_\nu \langle e', l' \rangle$ iff $e' \in \nu(e)$. Since ν partitions E, it is easy to see that $\sim_\nu$ is an equivalence relation. The semantic rule for the $\mathcal{V}$ modality is given in terms of the $\sim_\nu$ relation in the standard way for possible worlds semantics: $\langle M, \langle e, l \rangle \rangle \models \mathcal{V}\varphi$ iff $\langle M, \langle e', l' \rangle \rangle \models \varphi$ for all $\langle e', l' \rangle \in G_{Ag,Env}$ such that $\langle e, l \rangle \sim_\nu \langle e', l' \rangle$. As $\sim_\nu$ is an equivalence relation, the $\mathcal{V}$ modality has a logic of S5 [5]. In other words, formula schemas (1)-(5) are valid in $\mathcal{L}$:

$$\mathcal{V}(\varphi \Rightarrow \psi) \Rightarrow ((\mathcal{V}\varphi) \Rightarrow (\mathcal{V}\psi)) \tag{1}$$

$$\mathcal{V}\varphi \Rightarrow \neg\mathcal{V}\neg\varphi \tag{2}$$

$$\mathcal{V}\varphi \Rightarrow \varphi \tag{3}$$

$$\mathcal{V}\varphi \Rightarrow \mathcal{V}(\mathcal{V}\varphi) \tag{4}$$

$$\neg\mathcal{V}\varphi \Rightarrow \mathcal{V}\neg\mathcal{V}\varphi \tag{5}$$

We will omit the (by now standard) proof of this result — see, e.g., [5, pp58-59].

Formula schema (3) captures the first significant interaction between what is true and what is visible. However, we can also consider the converse of this implication:

$$\varphi \Rightarrow \mathcal{V}\varphi \tag{6}$$

This schema says that if φ is true of an environment, then φ is knowable. We can characterise this schema in terms of the environment's visibility function: formula schema (6) is valid in a model iff the visibility function of that model is transparent. Thus in transparent environments, visibility collapses to truth, since $\varphi \Leftrightarrow \mathcal{V}\varphi$ will be valid in such environments. In other words, everything true in a transparent environment is also visible, and *vice versa*. Note that we consider this a *helpful* property of environments — in the terminology of [13], such environments are *accessible*. Unfortunately, most environments do not enjoy this property.

4 Visibility and Perception

The fact that something is visible in an environment does not mean that an agent actually sees it. What an agent *does* see is determined by its sensors, which in our formal model are represented by the *see* function. In this section, we extend our logic by introducing a unary modal operator "$\mathcal{S}$", which is intended to allow us to represent the information that an agent sees. The intuitive meaning of a formula $\mathcal{S}\varphi$ is thus that the agent perceives the information φ. Note that, as with the $\mathcal{V}$ operator, the argument to $\mathcal{S}$ is a *proposition*, and *not* a term denoting an object.

In order to define the semantics of $\mathcal{S}$, we introduce a *perception accessibility relation* $\sim_s \subseteq G_{Ag,Env} \times G_{Ag,Env}$ as follows: $\langle e, l \rangle \sim_s \langle e', l' \rangle$ iff $see(\nu(e)) = see(\nu(e'))$. That is, $g \sim_s g'$ iff the agent receives the same percept when the system is in state g as it does in state g'. Again, it is straightforward to see that $\sim_s$ is an equivalence relation. Note that, for any of our models, it turns out that $\sim_\nu \subseteq \sim_s$.

The semantic rule for $\mathcal{S}$ is: $\langle M, \langle e, l \rangle \rangle \models \mathcal{S}\varphi$ iff $\langle M, \langle e', l' \rangle \rangle \models \varphi$ for all $\langle e', l' \rangle \in G_{Ag,Env}$ such that $\langle e, l \rangle \sim_s \langle e', l' \rangle$. As $\sim_s$ is an equivalence relation, $\mathcal{S}$ will also validate analogues of the S5 modal axioms KDT45:

$$\mathcal{S}(\varphi \Rightarrow \psi) \Rightarrow ((\mathcal{S}\varphi) \Rightarrow (\mathcal{S}\psi)) \tag{7}$$

$$\mathcal{S}\varphi \Rightarrow \neg\mathcal{S}\neg\varphi \tag{8}$$

$$\mathcal{S}\varphi \Rightarrow \varphi \tag{9}$$

$$\mathcal{S}\varphi \Rightarrow \mathcal{S}(\mathcal{S}\varphi) \tag{10}$$

$$\neg\mathcal{S}\varphi \Rightarrow \mathcal{S}\neg\mathcal{S}\varphi \tag{11}$$

It is worth asking whether these schemas are appropriate for a logic of perception. If we were attempting to develop a logic of *human* perception, then an S5 logic would *not* be acceptable. Human perception is often faulty, for example, thus rejecting schema (9). We would almost certainly reject (11), for similar reasons. However, our interpretation of $\mathcal{S}\varphi$ is that *the percept received by the agent carries the information* φ. Under *this* interpretation, an S5 logic seems appropriate.

We now turn to the relationship between $\mathcal{V}$ and $\mathcal{S}$. Given two unary modal operators, $\Box_1$ and $\Box_2$, the most important interactions between them can be summarised as follows:

$$\Box_1\varphi \underset{\Leftarrow}{\overset{\Rightarrow}{}} \Box_2\varphi \tag{$*$}$$

We use $(*)$ as the basis of our investigation of the relationship between $\mathcal{V}$ and $\mathcal{S}$. The most important interaction axiom says that if an agent sees φ, then φ must be visible.

It turns out that formula schema (12), which characterises this relationship, is valid — this follows from the fact that $\sim_s \subseteq \sim_\nu$.

$$\mathcal{S}\varphi \Rightarrow \mathcal{V}\varphi \tag{12}$$

Turning to the converse direction, the next interaction says that if φ is visible, then φ is seen by the agent — in other words, the agent sees everything visible.

$$\mathcal{V}\varphi \Rightarrow \mathcal{S}\varphi \tag{13}$$

Intuitively, this axiom characterises agents with "perfect" sensory apparatus, i.e., a *see* function that *never loses information*. Formally, we will say a perception function *see* is *perfect* iff it is an injection; otherwise we will say it is *lossy*. Lossy perception functions can map different visibility sets to the same percept, and hence, intuitively lose information. It turns out that formula schema (13) is valid in a model if the perception function of that model is perfect.

5 Perception and Knowledge

We now extend our language $\mathcal{L}$ by the addition of a unary modal operator $\mathcal{K}$. The intuitive meaning of a formula $\mathcal{K}\varphi$ is that the agent knows φ. In order to give a semantics to $\mathcal{K}$, we introduce a *knowledge accessibility relation* $\sim_k \subseteq G_{Ag,Env} \times G_{Ag,Env}$ in the by-now conventional way [5, p111]: $\langle e, l\rangle \sim_k \langle e', l'\rangle$ iff $l = l'$. As with $\sim_\nu$ and $\sim_s$, it is easy to see that $\sim_k$ is an equivalence relation. The semantic rule for $\mathcal{K}$ is as expected: $\langle M, \langle e, l\rangle\rangle \models \mathcal{K}\varphi$ iff $\langle M, \langle e', l'\rangle\rangle \models \varphi$ for all $\langle e', l'\rangle \in G_{Ag,Env}$ such that $\langle e, l\rangle \sim_k \langle e', l'\rangle$. Obviously, as with $\mathcal{V}$ and $\mathcal{S}$, the $\mathcal{K}$ modality validates analogues of the modal axioms KDT45.

$$\mathcal{K}(\varphi \Rightarrow \psi) \Rightarrow ((\mathcal{K}\varphi) \Rightarrow (\mathcal{K}\psi)) \tag{14}$$
$$\mathcal{K}\varphi \Rightarrow \neg\mathcal{K}\neg\varphi \tag{15}$$
$$\mathcal{K}\varphi \Rightarrow \varphi \tag{16}$$
$$\mathcal{K}\varphi \Rightarrow \mathcal{K}(\mathcal{K}\varphi) \tag{17}$$
$$\neg\mathcal{K}\varphi \Rightarrow \mathcal{K}\neg\mathcal{K}\varphi \tag{18}$$

We now turn to the relationship between what an agent perceives and what it knows. As with the relationship between $\mathcal{S}$ and $\mathcal{V}$, the main interactions of interest are captured in $(*)$. The first interaction we consider states that when an agent sees something, it knows it.

$$\mathcal{S}\varphi \Rightarrow \mathcal{K}\varphi \tag{19}$$

Intuitively, this property will be true of an agent if its next state function distinguishes between every different percept received. If a next state function has this property, then intuitively, it never loses information from the percepts. We say a next state function is *complete* if it distinguishes between every different percept. Formally, a next state function *next* is *complete* iff $next(l, \rho) = next(l', \rho')$ implies $\rho = \rho'$. Formula schema (19) is valid in a model iff the next state function of that model is complete.

Turning to the converse direction, we might expect the following schema to be valid:

$$\mathcal{K}\varphi \Rightarrow \mathcal{S}\varphi \tag{20}$$

While this schema is satisfiable, it is not valid. To understand what kinds of agents validate this schema, imagine an agent with a next state function that chooses the next state *solely* on the basis of it current state. Let us say that an agent is *local* if it has this property. Formally, an agent's next-state function is local iff $next(l, \rho) = next(l', \rho)$ for all local states $l, l' \in L$, and percepts $\rho \in P$. It is not hard to see that formula schema (20) is valid in a model if the next state function of the agent in this model is local.

6 Systems of $\mathcal{VSK}$ Logic

The preceding sections identified the key interactions that may hold between what is true, visible, seen, and known. In this section, we consider *systems* of $\mathcal{VSK}$ logic, by which we mean possible *combinations* of interactions that could hold for any given agent-environment system. To illustrate, consider the class of systems in which: (i) the environment is not transparent; (ii) the agent's perception function is perfect; and (iii) the agent's next state function is neither complete nor local. In this class of models, the formula schemas (3), (12), and (13) are valid. These formula schemas can be understood as characterising a class of agent-environment systems — those in which the environment is not transparent, the agent's perception function is perfect, and the agent's next state function is neither complete nor local. In this way, by systematically considering the possible combinations of $\mathcal{VSK}$ formula schemas, we obtain a classification scheme for agent-environment systems. As the basis of this scheme, we consider only interaction schemas with the following form.

$$\Box_1\varphi \overset{\Rightarrow}{\underset{\Leftarrow}{}} \Box_2\varphi$$

Given the three $\mathcal{VSK}$ modalities there are six such interaction schemas: (6), (3), (13), (12), (19), and (20). This in turn suggests there should be 64 distinct $\mathcal{VSK}$ systems. However, as (3) and (12) are valid in all $\mathcal{VSK}$ systems, there are in fact only 16 distinct systems, summarised in Table 1.

In systems $\mathcal{VSK}$-8 to $\mathcal{VSK}$-15 inclusive, visibility and truth are equivalent, in that everything true is also visible. These systems are characterised by transparent visibility relations. Formally, the schema $\varphi \Leftrightarrow \mathcal{V}\varphi$ is a valid formula in systems $\mathcal{VSK}$-8 to $\mathcal{VSK}$-15. The $\mathcal{V}$ modality is redundant in such systems.

In systems $\mathcal{VSK}$-4 to $\mathcal{VSK}$-7 and $\mathcal{VSK}$-12 to $\mathcal{VSK}$-15, everything visible is seen, and everything seen is visible. Visibility and perception are thus equivalent: the formula schema $\mathcal{V}\varphi \Leftrightarrow \mathcal{S}\varphi$ is valid in such systems. Hence one of the modalities $\mathcal{V}$ or $\mathcal{S}$ is redundant in systems $\mathcal{VSK}$-4 to $\mathcal{VSK}$-7 and $\mathcal{VSK}$-12 to $\mathcal{VSK}$-15. Models for these systems are characterised by agents with perfect perception (*see*) functions.

System Name	Formula Schemas (6) $\varphi \Rightarrow \mathcal{V}\varphi$	(3) $\mathcal{V}\varphi \Rightarrow \varphi$	(13) $\mathcal{V}\varphi \Rightarrow \mathcal{S}\varphi$	(12) $\mathcal{S}\varphi \Rightarrow \mathcal{V}\varphi$	(19) $\mathcal{S}\varphi \Rightarrow \mathcal{K}\varphi$	(20) $\mathcal{K}\varphi \Rightarrow \mathcal{S}\varphi$
$\mathcal{VSK}$-0		$\times$		$\times$		
$\mathcal{VSK}$-1		$\times$		$\times$		$\times$
$\mathcal{VSK}$-2		$\times$		$\times$	$\times$	
$\mathcal{VSK}$-3		$\times$		$\times$	$\times$	$\times$
$\mathcal{VSK}$-4		$\times$	$\times$	$\times$		
$\mathcal{VSK}$-5		$\times$	$\times$	$\times$		$\times$
$\mathcal{VSK}$-6		$\times$	$\times$	$\times$	$\times$	
$\mathcal{VSK}$-7		$\times$	$\times$	$\times$	$\times$	$\times$
$\mathcal{VSK}$-8	$\times$	$\times$		$\times$		
$\mathcal{VSK}$-9	$\times$	$\times$		$\times$		$\times$
$\mathcal{VSK}$-10	$\times$	$\times$		$\times$	$\times$	
$\mathcal{VSK}$-11	$\times$	$\times$		$\times$	$\times$	$\times$
$\mathcal{VSK}$-12	$\times$	$\times$	$\times$	$\times$		
$\mathcal{VSK}$-13	$\times$	$\times$	$\times$	$\times$		$\times$
$\mathcal{VSK}$-14	$\times$	$\times$	$\times$	$\times$	$\times$	
$\mathcal{VSK}$-15	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

Table 1. The sixteen possible $\mathcal{VSK}$ systems. A cross ($\times$) indicates that the schema is valid in the corresponding system; all systems include (3) and (12).

In systems $\mathcal{VSK}$-3, $\mathcal{VSK}$-7, $\mathcal{VSK}$-11, and $\mathcal{VSK}$-15, knowledge and perception are equivalent: an agent knows everything it sees, and sees everything it knows. In these systems, $\mathcal{S}\varphi \Leftrightarrow \mathcal{K}\varphi$ is valid. Models of such systems are characterised by complete, local next state functions.

In systems $\mathcal{VSK}$-12 to $\mathcal{VSK}$-15, we find that truth, visibility, and perception are equivalent: the schema $\varphi \Leftrightarrow \mathcal{V}\varphi \Leftrightarrow \mathcal{S}\varphi$ is valid. In such systems, the $\mathcal{V}$ and $\mathcal{S}$ modalities are redundant.

An analysis of individual $\mathcal{VSK}$ systems identifies a number of interesting properties, but space limitations prevents such an analysis here. We simply note that in system $\mathcal{VSK}$-15, the formula schema $\varphi \Leftrightarrow \mathcal{V}\varphi \Leftrightarrow \mathcal{S}\varphi \Leftrightarrow \mathcal{K}\varphi$ is valid, and hence all three modalities $\mathcal{V}$, $\mathcal{S}$, and $\mathcal{K}$ are redundant. System $\mathcal{VSK}$-15 thus collapses to propositional logic.

7 Related Work

Since the mid 1980s, Halpern and colleagues have used modal epistemic logic for reasoning about multi-agent systems [5]. 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. Comparatively little effort has been devoted to characterising "architectural" properties of agents. The only obvious examples are the properties of no learning, perfect recall, and so on [5, pp281–307].

In their "situated automata" paradigm, Kaelbling and Rosenschein directly synthesised agents (in fact, digital circuits) from epistemic specifications of these agents [12]. 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 [2].

Many other formalisms for reasoning about intelligent agents and multi-agent systems have been proposed over the past decade [16]. Following the pioneering work of Moore on the interaction between knowledge and action [9], 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 [4], and the ongoing work of Rao-Georgeff on the belief-desire-intention (BDI) model of agency [10]. The emphasis in this work has been more on axiomatic characterisations of architectural properties; for example, in [11], 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.

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 [9], the goal of such work is typically to develop representations of sensing actions that can be used in planning algorithms [1]. An example is [14], in which Scherl and Levesque develop a representation of sensing actions in the situation calculus [8]. 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.

Finally, it is worth noting that there is now a growing body of work addressing the abstract logical properties of multi-modal logics, of which $\mathcal{VSK}$ is an example [3]. Lomuscio and Ryan, for example, investigates axiomatizations of multi-agent epistemic logic (epistemic logics with multiple $\mathcal{K}$ operators) [7]. The work in this paper can clearly benefit from such work.

8 Conclusions

In this paper, we have presented a formalism that allows us to represent several key aspects of the relationship between an agent and the environment in which it is situated. Specifically, it allows us to distinguish between what is true of an environment and what is visible, or knowable about it; what is visible of an environment and what an agent actually perceives of it; and what an agent perceives of an environment and actually knows of it. Previous formalisms do not permit us to make such distinctions.

For future work, a number of obvious issues present themselves:

- *Completeness*.
  First, completeness results for the formalism would be desirable: multi-modal logics are a burgeoning area of research, for which general completeness results are beginning to emerge.
- *Multi-agent extensions*.
  Another issue is extending the formalism to the multi-agent domain. It would be interesting to investigate such interactions as $\mathcal{K}_i \mathcal{V}_j \varphi$ (agent i knows that φ is visible to agent j).
- *Temporal extensions*.
  The emphasis in this work has been on classifying instananeous relationships in $\mathcal{VSK}$ logic. Much work remains to be done in considering the temporal extensions to the logic, in much the same way that epistemic logic is extended into the temporal dimension in [15].
- *Knowledge-based programs*.
  The relationship between $\mathcal{VSK}$ logic and knowledge-based programs [5, Chapter 7] would also be an interesting area of future work: $\mathcal{VSK}$ logic has something to say about when such programs are implementable.

References

1. J. F. Allen, J. Hendler, and A. Tate, editors. *Readings in Planning*. Morgan Kaufmann Publishers: San Mateo, CA, 1990.
2. 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.
3. L. Catach. Normal multimodal logics. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-88)*, pages 491–495, St. Paul, MN, 1988.
4. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
5. R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. The MIT Press: Cambridge, MA, 1995.
6. M. R. Genesereth and N. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers: San Mateo, CA, 1987.
7. Alessio Lomuscio and Mark Ryan. A spectrum of modes of knowledge sharing between agents. In N.R. Jennings and Y. Lespérance, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin, 2000. In this volume.
8. 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.
9. 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.
10. A. S. Rao and M. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–344, 1998.

11. 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.
12. 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.
13. S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, 1995.
14. 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.
15. M. Wooldridge, C. Dixon, and M. Fisher. A tableau-based proof method for temporal logics of knowledge and belief. *Journal of Applied Non-Classical Logics*, 8(3):225–258, 1998.
16. M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.