# Multi-Agent $\mathcal{VSK}$ Logic

Michael Wooldridge* and Alessio Lomuscio†

* Department of Computer Science
University of Liverpool
Liverpool L69 7ZF, United Kingdom
M.J.Wooldridge@csc.liv.ac.uk

† Department of Computing
Imperial College of Science, Technology and Medicine
London SW7 2BZ, United Kingdom
A.Lomuscio@doc.ic.ac.uk

**Abstract.** We present a formalism for reasoning about the information properties of multi-agent systems. Multi-agent $\mathcal{VSK}$ logic allows us to represent what is *objectively true* of some environment, what is *visible*, or *accessible* of the environment to individual agents, what these agents actually *perceive*, and finally, what the agents actually *know* about the environment. The semantics of the logic are given in terms of a general model of multi-agent systems, closely related to the interpreted systems of epistemic logic. After introducing the logic and establishing its relationship to the formal model of multi-agent systems, we systematically investigate a number of possible interaction axioms, and characterise these axioms in terms of the properties of agents that they correspond to. Finally, we illustrate the use of the logic through a case study, and discuss issues for future work.

## 1 Introduction

Consider the following scenario:

> A number of autonomous mobile robots are working in a factory, collecting and moving various goods around. All robots are equipped with sonars, which enable them to detect obstacles. To ensure that potentially costly collisions are avoided, a number of crash-avoidance techniques are used. First, all robots adhere to a convention that, if they detect a potential collision, they must take evasive action either when they detect that other agents have right of way or when they know that regardless of the convention of the right of way this is the only way to avoid a collision. Second, a "supervisor" agent *C* is installed in the factory, which monitors all data feeds from sonars. In the event of an impending collision, this agent is able to step in and override the control systems of individual agents. At some time, two robots, *A* and *B*, are moving towards each other in a narrow corridor; robot *A* has the right of way. Robot *B*'s sonar is faulty, and as a result, *B* fails to notice the potential collision and does not give way to robot *A*. Robot *A*, using its sonar, detects the presence of robot *B*.

> Robot $A$ recognises that $B$ has not taken evasive action when it should have done, and reasons that $B$ must be faulty; as a consequence, it takes additional evasive action. Meanwhile, the supervisor agent $C$, observing the scenario, also deduces that $B$ must be faulty, and as a consequence shuts $B$ down.

The aim of this scenario is not to suggest an architecture for multi-agent robotics, but to illustrate the utility of reasoning about the information that agents can and do perceive, their knowledge about their environment, and the actions that they perform. We argue that the ability to perform such reasoning will be of great value if autonomous agents are to be successfully deployed.

In this paper, we develop a formalism that will allow us to represent and reason about such aspects of multi-agent systems. We present *multi-agent $\mathcal{VSK}$ logic*, a multi-agent extension of $\mathcal{VSK}$ logic [9]. This logic allows us to represent what is *objectively true* of an environment, what is *visible*, or *knowable* about the environment to individual agents within it, what agents *perceive* of their environment, and finally, what agents actually *know* about their environment. Syntactically, $\mathcal{VSK}$ logic is a propositional multi-modal logic, containing three sets of indexed unary modal operators "$\mathcal{V}_i$", "$\mathcal{S}_i$", and "$\mathcal{K}_i$", one for each agent $i$. A formula $\mathcal{V}_i\varphi$ means that the information $\varphi$ is accessible to agent $i$; $\mathcal{S}_i\varphi$ means that agent $i$ perceives information $\varphi$; and $\mathcal{K}_i\varphi$ means that agent $i$ knows $\varphi$.

An important feature of multi-agent $\mathcal{VSK}$ logic is that its semantics are given with respect to a general model of agents and their environments. We are able to characterise possible axioms of multi-agent $\mathcal{VSK}$ logic with respect to this semantic model. Consider, for example, the $\mathcal{VSK}$ formula $\mathcal{V}_i\varphi \Rightarrow \mathcal{S}_j\mathcal{V}_i\varphi$, which says that if information $\varphi$ is accessible to agent $i$, then agent $j$ sees (perceives) that $\varphi$ is accessible to $i$. Intuitively, this formula says that agent $j$ is able to see at least as much as agent $i$; we are able to show this formally by proving correspondence results with respect to a semantic description of agents and environments, as well as the Kripke frames they generate.

The remainder of this paper is structured as follows. We begin in section 2 by introducing the semantic framework that underpins multi-agent $\mathcal{VSK}$ logic. We then formally introduce the syntax and semantics of $\mathcal{VSK}$ logic in section 3, and in particular, we show how the semantics of the logic relate to the formal model of multi-agent systems introduced in section 2. In section 4, we discuss and formally characterise various *interaction axioms* of $\mathcal{VSK}$ logic. In section 5, we return to the case study presented above, and show how we can use multi-agent $\mathcal{VSK}$ logic to capture and reason about

Finally, in section 6, we present some conclusions.

## 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 [2, pp103–107].

A *multi-agent $\mathcal{VSK}$ system* is assumed to be comprised of a collection $Ag_1, \ldots, Ag_n$ of *agents*, together with an *environment*. We formally define environments below, but for the moment, it is assumed that an environment can be in any of a set $E$ of *instantaneous states*. We adopt a quite 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.

Formally, an agent $Ag_i$ is a tuple $Ag_i = \langle L_i, Act_i, see_i, do_i, \tau_i, \mathbf{l}_i \rangle$, where:

- $L_i = \{l_i^1, l_i^2, \ldots\}$ is a set of *instantaneous local states* for agent $i$.
- $Act_i = \{\alpha_i^1, \alpha_i^2, \ldots\}$ is a set of *actions* for agent $i$.
- $see_i : 2^E \rightarrow Perc_i$ is the *perception function* for agent $i$, mapping sets of environment states (*visibility sets*) to *percepts* for agent $i$.
  Elements of the set $Perc_i$ will be denoted by $\rho_i^1, \rho_i^2, \ldots$ and so on. If $see_i$ is an injection into $Perc_i$ then we say that $see_i$ is *perfect*, otherwise we say it is *lossy*.
- $do_i : L_i \rightarrow Act_i$ is the *action selection function* for agent $i$, mapping local states to actions available to agent $i$.
- $\tau_i : L_i \times Perc_i \rightarrow L_i$ is the *state transformer function* for agent $i$.
  We say $\tau_i$ is *complete* if for any
  $$g = (e, \tau_1(l_1, \rho_1), \ldots, \tau_n(l_n, \rho_n))$$
  and
  $$g' = (e', \tau_1(l_1', \rho_1'), \ldots, \tau_n(l_n', \rho_n'))$$
  we have that
  $$\tau_i(l_i, \rho_i) = \tau_i(l_i', \rho_i') \quad \text{implies} \quad \rho_i = \rho_i'.$$
  We say $\tau_i$ is *local* if for any
  $$g = (e, \tau_1(l_1, \rho_1), \ldots, \tau_n(l_n, \rho_n))$$
  and
  $$g' = (e', \tau_1(l_1', \rho_1'), \ldots, \tau_n(l_n', \rho_n'))$$
  we have that
  $$\tau_i(l_i, \rho_i) = \tau_i(l_i', \rho_i).$$
  We say that an agent has *perfect recall* if the function $\tau_i$ is an injection.
- $\mathbf{l}_i \in L$ is the *initial state* for agent $i$.

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. We say that an agent has *perfect recall* of its history if it changes its local state at every tick of the clock (cf. [2, pp128–131]).

Following [2], we use the term "environment" to denote all the components of a system external to the agents that occupy it. 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 tuple containing a set of possible *instantaneous states*, a *visibility function* for each agent, which characterises the information available to an agent in every environment state, a *state transformer* function, which characterises the effects that an agent's actions have on the environment, and, finally, an *initial state*.

Formally, an environment *Env* is a tuple

$$Env = \langle E, vis_1, \ldots, vis_n, \tau_e, e_0 \rangle$$

where:

- $E = \{e_1, e_2, \ldots\}$ is a *set of instantaneous local states* for the environment.
- $vis_i : E \to 2^E$ is the *visibility function of agent i*. It is assumed that $vis_i$ partitions $E$ into mutually disjoint sets and that $e \in vis_i(e)$, for any $e \in E$. Elements of the codomain of the function *vis* are called *visibility sets*. We say that $vis_i$ is *transparent* if for any $e \in E$ we have that $vis_i(e) = \{e\}$.
- $\tau_e : E \times Act_1 \times \cdots \times Act_n \to 2^E$ is a total *state transformer function* for the environment (cf. [2, p154]), which maps environment states and tuples of actions, one for each agent, to the set of environment states that could result from the performance of these actions in this state.
- $e_0 \in E$ 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., [2]). The use of the visibility function, however, requires some explanation. Before we do this, let us define the concept of global state. The *global states* $G = \{g, g', \ldots\}$ of a $\mathcal{VSK}$ system are a subset of $E \times L_1 \times \cdots \times L_n$.

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 [6]. Intuitively, not all the information in an environment state is in general accessible to an agent. So, in a global state $g = (e, l_1, \ldots, l_n)$, $vis_i(e) = \{e, e', e''\}$ represents the fact that the environment states $e, e', e''$ are indistinguishable to agent $i$ from $e$. This is so regardless of agent $i$'s efforts in performing the observation — it represents the maximum amount of information that is in principle available to $i$ when observing state $e$. The concept of transparency, as defined above, 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".

A *multi-agent* $\mathcal{VSK}$ *system* is a structure $S = \langle Env, Ag_1, \ldots, Ag_n \rangle$, where *Env* is an environment, and $Ag_1, \ldots, Ag_n$ are agents. 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 a $\mathcal{VSK}$ system can be summarised as follows. Each agent $i$ starts in state $\mathbf{l}_i$, the environment starts in state $e_0$. At this point every agent $i$ "synchronises" with the environment by performing an initial observation through the visibility function $vis_i$, and generates a percept $\rho_i^0 = see_i(vis_i(e_0))$. The internal state of the agent is then updated, and becomes $\tau_i(\mathbf{l}_i, \rho_i^0)$. The synchronisation phase is now over and the system starts its run from the initial state $g_0 = (e_0, \tau_1(\mathbf{l}_1, \rho_1^0), \ldots, \tau_n(\mathbf{l}_n, \rho_n^0))$. An action $\alpha_i^0 = do(\tau_i(\mathbf{l}_i, \rho_i^0))$ is selected and performed by each agent $i$ on the environment, whose state is updated into $e_1 = \tau_e(e_0, \alpha_1^0, \ldots, \alpha_n^0)$. Each agent enters another cycle, and so on.

A *run* of a system is thus a (possibly infinite) sequence of global states. A sequence $(g_0, g_1, g_2, \ldots)$ over $G$ represents a run of a system $\langle Env, Ag_1, \ldots, Ag_n \rangle$ iff

– $g_0 = (e_0, \tau_1(\mathbf{l}_1, see_1(vis_1(e_0))), \ldots, \tau_n(\mathbf{l}_n, see_n(vis_n(e_0))))$, and
– for all $u$, if $g_u = (e, l_1, \ldots, l_n)$ and $g_{u+1} = (e', l'_1, \ldots, l'_n)$ then:

$$e' \in \tau_e(e_u, \alpha_1, \ldots, \alpha_n) \quad \text{and}$$
$$l'_i = \tau_i(l_i, see_i(vis_i(e')))$$

where $\alpha_i = do_i(l_i)$.

Given a multi-agent $\mathcal{VSK}$ system $S = \langle Env, Ag_1, \ldots, Ag_n \rangle$, we say $G_S \subseteq G$ is the set of global states *generated* by $S$ if $g \in G_S$ occurs in a run of $S$.

## 3 Multi-Agent $\mathcal{VSK}$ Logic

We now introduce a language $\mathcal{L}$, which will enable us to represent the information properties of multi-agent $\mathcal{VSK}$ 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 to the agents within it, then what these agents *perceive* of the system, and finally, what each agent *knows* of the system. $\mathcal{L}$ is a propositional multi-modal language, containing three sets of indexed unary modal operators, for visibility, perception, and knowledge respectively. Given a set $P$ of propositional atoms, the language $\mathcal{L}$ of $\mathcal{VSK}$ logic is defined by the following BNF grammar:

$$\begin{aligned} \langle ag \rangle &::= 1 \mid \cdots \mid n \\ \langle wff \rangle &::= \textbf{true} \mid \text{any element of } P \mid \neg\langle wff \rangle \mid \langle wff \rangle \wedge \langle wff \rangle \\ &\mid \; \mathcal{V}_{\langle ag \rangle}\langle wff \rangle \mid \mathcal{S}_{\langle ag \rangle}\langle wff \rangle \mid \mathcal{K}_{\langle ag \rangle}\langle wff \rangle \end{aligned}$$

The modal operator "$\mathcal{V}_i$" will allow us to represent the information that is instantaneously visible or knowable about the state of the system to agent $i$. Thus suppose the formula $\mathcal{V}_i\varphi$ is true in some state $g \in G$. The intended interpretation of this formula is that the property $\varphi$ is *accessible* to agent $i$ when the system is in state $g$. This means that not only $\varphi$ is true of the environment, but agent $i$, if it was equipped with suitable sensor apparatus, would be able to perceive $\varphi$. If $\neg\mathcal{V}_i\varphi$ were true in some state, then no matter how good agent $i$'s sensor apparatus was, it would be unable to perceive $\varphi$.

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

Finally, we can represent the *knowledge* possessed by agents within a system. We represent agent $i$'s knowledge by means of a modal operator "$\mathcal{K}_i$". In line with the tradition that started with Hintikka [4], we write $\mathcal{K}_i\varphi$ to represent the fact that agent $i$ has knowledge of the formula represented by $\varphi$. Our model of knowledge is that popularised by Halpern and colleagues [2]: agent $i$ is said to know $\varphi$ when in local state

$l$ if $\varphi$ is guaranteed to be true whenever $i$ is in state $l$. As with visibility and perception, knowledge is an *external* notion — an agent is said to know $\varphi$ if an impartial, omniscient observer would say that the agent's state carried the information $\varphi$.

We now proceed to interpret our formal language. We do so with respect to the equivalence Kripke frames *generated* (see [2]) by $\mathcal{VSK}$ systems. Given a $\mathcal{VSK}$ system $S = \langle Env, Ag_1, \ldots, Ag_n \rangle$, the Kripke frame

$$F_S = \langle W, \sim_1^\nu, \sim_1^s, \sim_1^k, \ldots, \sim_n^\nu, \sim_n^s, \sim_n^k \rangle$$

generated by $S$ is defined as follows:

- $W = G_S$ (recall that $G_S$ is the set of global states reachable by system $S$),
- For every $i = 1, \ldots, n$, the relation $\sim_i^\nu \subseteq W \times W$ is defined by: $(e, l_1, \ldots, l_n) \sim_i^\nu (e', l'_1, \ldots, l'_n)$ if $e' \in vis_i(e)$,
- For every $i = 1, \ldots, n$, the relation $\sim_i^s \subseteq W \times W$ is defined by: $(e, l_1, \ldots, l_n) \sim_i^s (e', l'_1, \ldots, l'_n)$ if $see_i(vis_i(e)) = see_i(vis_i(e'))$,
- For every $i = 1, \ldots, n$, the relation $\sim_i^k \subseteq W \times W$ is defined by: $(e, l_1, \ldots, l_n) \sim_i^k (e', l'_1, \ldots, l'_n)$ if $l_i = l'_i$.

The class of frames generated by a $\mathcal{VSK}$ system $\mathcal{S}$ will be denoted by $\mathcal{F}_\mathcal{S}$. As might be expected, all frames generated by systems in $\mathcal{S}$ are equivalence frames.

**Lemma 1.** *Every frame $F \in \mathcal{F}_\mathcal{S}$ is an equivalence frame, i.e., all the relations in $F$ are equivalence relations.*

We have now 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 — we refer the reader to [5, 3] for a detailed exposition of the subject. Following [2] and [7], we define the concepts of truth and validity on Kripke models that are *generated* by $\mathcal{VSK}$ systems.

Given an interpretation $\pi : W \to 2^P$, we say that a formula $\varphi \in \mathcal{L}$ 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 $\pi$ is such that $M_S \models_g \varphi$. The propositional connectives are assumed to be interpreted as usual, and the modal operators $\mathcal{V}_i$, $\mathcal{S}_i$, and $\mathcal{K}_i$ are assumed to be interpreted in the standard way (see for example [5]) by means of the equivalence relations $\sim_i^\nu$, $\sim_i^s$, and $\sim_i^k$ 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. A formula $\varphi \in \mathcal{L}$ is valid on a class $\mathcal{S}$ of $\mathcal{VSK}$ systems if for any system $S \in \mathcal{S}$, we have that $F_S \models \varphi$.

## 4 Interaction Axioms in Multi-Agent $\mathcal{VSK}$ Logic

In this section we will study some basic interaction axioms that can be specified within $\mathcal{VSK}$ logic. Interaction axioms are formulas in which different modalities are present; they specify a form of "binding" between the attitudes corresponding to the modal operators.

| Axiom | $\mathcal{VSK}$ Class |
|---|---|
| $\mathcal{V}_i\varphi \Rightarrow \varphi$ | none (valid in all systems) |
| $\varphi \Rightarrow \mathcal{V}_i\varphi$ | $vis_i$ is transparent |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{V}_i\varphi$ | none (valid in all systems) |
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{S}_i\varphi$ | $see_i$ is perfect |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{S}_i\varphi$ | $\tau_i$ has perfect recall |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{K}_i\varphi$ | $\tau_i$ is local |

**Table 1.** Single-agent interaction axioms in $\mathcal{VSK}$ logic.

| Axiom | Kripke Condition | $\mathcal{VSK}$ Condition |
|---|---|---|
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{V}_j\varphi$ | $\sim_j^v \subseteq \sim_i^v$ | $vis_j(e) \subseteq vis_i(e)$ |
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{S}_j\varphi$ | $\sim_j^s \subseteq \sim_i^v$ | $sv_j(e) = sv_j(e') \rightarrow vis_i(e) = vis_i(e')$ |
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{K}_j\varphi$ | $\sim_j^k \subseteq \sim_i^v$ | $l_j = l_j' \rightarrow vis_i(e) = vis_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{V}_j\varphi$ | $\sim_j^v \subseteq \sim_i^s$ | $vis_j(e) = vis_j(e') \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{S}_j\varphi$ | $\sim_j^s \subseteq \sim_i^s$ | $sv_j(e) = sv_j(e') \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{K}_j\varphi$ | $\sim_j^k \subseteq \sim_i^s$ | $l_j = l_j' \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{V}_j\varphi$ | $\sim_j^v \subseteq \sim_i^k$ | $vis_j(e) = vis_j(e') \rightarrow l_i = l_i'$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{S}_j\varphi$ | $\sim_j^s \subseteq \sim_i^k$ | $sv_j(e) = sv_i(e') \rightarrow l_i = l_i'$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{K}_j\varphi$ | $\sim_j^k \subseteq \sim_i^k$ | $l_j = l_j' \rightarrow l_i = l_i'$ |

**Table 2.** Some multi-agent interaction axioms in multi-agent $\mathcal{VSK}$ logic. Note that in the table the function $sv_i : E \rightarrow Perc_i$ stands for $see_i \circ vis_i$.

Note that, in previous work, we have studied and given semantic characterisations for *single-agent* interaction axioms (i.e., axioms in a $\mathcal{VSK}$ logic where there is only one $\mathcal{V}$ operator, only one $\mathcal{S}$ operator, and only one $\mathcal{K}$ operator) [9]. For example, we were able to show that the axiom schema $\mathcal{V}\varphi \Rightarrow \mathcal{S}\varphi$ characterised a particular property of an agent's perception function: namely, that it was *perfect*, in the sense that we defined in section 2. We summarise these results in table 1.

In this paper we analyse some *multi-agent* interaction axioms. The most obvious form that these interaction axioms may have is the following:

$$\boxdot_i\varphi \Rightarrow \boxdot_j\varphi \qquad \text{where } \boxdot_i \in \{\mathcal{S}_i, \mathcal{V}_i, \mathcal{K}_i\}, \boxdot_j \in \{\mathcal{S}_j, \mathcal{V}_j, \mathcal{K}_j\}. \tag{1}$$

If we assume $i \neq j$ (the case $i = j$ was dealt with in [9]), Axiom (1) generates nine possible interaction axioms in total, as summarised in table 2. The second column of table 1 gives the conditions on Kripke models that correspond (in the sense of [1]) to the axiom. The third column gives the first-order condition on $\mathcal{VSK}$ systems that corresponds to the interaction axioms. (Note that in these conditions each variable is assumed to be universally quantified: for example, the third axiom $\mathcal{V}_i\varphi \Rightarrow \mathcal{K}_j\varphi$ corresponds to systems $S$ in which for all $g = (e, l_1, \ldots, l_n)$ and $g' = (e', l_1', \ldots, l_n')$, we have that $l_j = l_j'$ implies $vis_i(e) = vis_i(e')$.)

We begin our analysis with the schema which says that if $\varphi$ is visible to $i$, then $\varphi$ is visible to $j$.

$$\mathcal{V}_i\varphi \Rightarrow \mathcal{V}_j\varphi \tag{2}$$

This axiom says that everything visible to $i$ is also visible to $j$. Note that the first-order condition corresponding to Axiom 2 implies that at least as much information is accessible to agent $j$ as agent $i$.

$$\mathcal{V}_i\varphi \Rightarrow \mathcal{S}_j\varphi \tag{3}$$

Axiom (3) says that $j$ sees everything visible to $i$. It is easy to see that in systems that validate this schema, since $j$ sees everything $i$ sees, it must be that everything visible to $i$ is also visible to $j$. In other words, $\mathcal{VSK}$ systems that validate Axiom (3) will also validate (2).

$$\mathcal{V}_i\varphi \Rightarrow \mathcal{K}_j\varphi \tag{4}$$

Axiom (4) says that everything visible to $i$ is known to $j$.

$$\mathcal{S}_i\varphi \Rightarrow \mathcal{V}_j\varphi \tag{5}$$

Axiom (5) says that everything $i$ sees is visible to $j$. Intuitively, this means that the percepts $i$ receives are part of the environment that is visible to $j$.

$$\mathcal{S}_i\varphi \Rightarrow \mathcal{S}_j\varphi \tag{6}$$

Axiom (6) says that $j$ sees everything $i$ sees. Since we know from [9] that any system $S$ validates the axiom $\mathcal{S}_j\varphi \Rightarrow \mathcal{V}_j\varphi$, it follows that any $\mathcal{VSK}$ system validating Axiom (6) will also validate Axiom (5). Note that from table 2, it follows that

$$|see_j(vis_j(E))| \leq |see_i(vis_i(E))|$$

So, since agent $i$ has more perception states at its disposal than agent $j$, it has a finer grain of perception.

$$\mathcal{S}_i\varphi \Rightarrow \mathcal{K}_j\varphi \tag{7}$$

Axiom (7) says that if $i$ sees $\varphi$ then $j$ knows $\varphi$; in other words, $j$ knows everything that $i$ sees.

$$\mathcal{K}_i\varphi \Rightarrow \mathcal{V}_j\varphi \tag{8}$$

Axiom (8) says that if $i$ knows $\varphi$, then $\varphi$ is visible to $j$. Intuitively, this means that $i$'s local state is visible to $j$. Axiom (8) thus says that entity $j$ has "read access" to the state of another entity $i$.

$$\mathcal{K}_i\varphi \Rightarrow \mathcal{S}_j\varphi \tag{9}$$

| Axiom | Kripke Condition | $\mathcal{VSK}$ Condition |
|---|---|---|
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{V}_j\mathcal{V}_i\varphi$ | $\sim_j^v \subseteq \sim_i^v$ | $vis_j(e) \subseteq vis_i(e)$ |
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{S}_j\mathcal{V}_i\varphi$ | $\sim_j^s \subseteq \sim_i^v$ | $sv_j(e) = sv_j(e') \rightarrow vis_i(e) = vis_i(e')$ |
| $\mathcal{V}_i\varphi \Rightarrow \mathcal{K}_j\mathcal{V}_i\varphi$ | $\sim_j^k \subseteq \sim_i^v$ | $l_j = l'_j \rightarrow vis_i(e) = vis_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{V}_j\mathcal{S}_i\varphi$ | $\sim_j^v \subseteq \sim_i^s$ | $vis_j(e) = vis_j(e') \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{S}_j\mathcal{S}_i\varphi$ | $\sim_j^s \subseteq \sim_i^s$ | $sv_j(e) = sv_j(e') \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{S}_i\varphi \Rightarrow \mathcal{K}_j\mathcal{S}_i\varphi$ | $\sim_j^k \subseteq \sim_i^s$ | $l_j = l'_j \rightarrow sv_i(e) = sv_i(e')$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{V}_j\mathcal{K}_i\varphi$ | $\sim_j^v \subseteq \sim_i^k$ | $vis_j(e) = vis_j(e') \rightarrow l_i = l'_i$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{S}_j\mathcal{K}_i\varphi$ | $\sim_j^s \subseteq \sim_i^k$ | $sv_j(e) = sv_i(e') \rightarrow l_i = l'_i$ |
| $\mathcal{K}_i\varphi \Rightarrow \mathcal{K}_j\mathcal{K}_i\varphi$ | $\sim_j^k \subseteq \sim_i^k$ | $l_j = l'_j \rightarrow l_i = l'_i$ |

**Table 3.** Other interaction axioms in multi-agent $\mathcal{VSK}$ logic.

Axiom (9) captures a more general case than that of (8), where entity *j* not only has read access to the state of *i*, but that it actually does read this state. Note that any system that validates (9) will also validate (8).

$$\mathcal{K}_i\varphi \Rightarrow \mathcal{K}_j\varphi \tag{10}$$

This final schema says that *j* knows everything that *i* knows. Note that from the corresponding condition on $\mathcal{VSK}$ systems in table 2, it follows that

$$|L_j| \leq |L_i|$$

So, since agent *i* has more local states, it has a finer grain of knowledge than agent *j*. If we also have the converse of (10), then we would have $\mathcal{K}_i\varphi \Leftrightarrow \mathcal{K}_j\varphi$ as valid; an obvious interpretation of this schema would be that *i* and *j* had the *same* state.

All these considerations lead us to the following:

**Theorem 1.** *For any axiom $\psi$ of table 2 and any $\mathcal{VSK}$ system $S$ we have that the following are equivalent:*

1. *The system $S$ validates $\psi$, i.e., $S \models \psi$;*
2. *The generated frame $F_S$ satisfies the corresponding Kripke condition $R_\psi$;*
3. *The system $S$ satisfies the corresponding $\mathcal{VSK}$ condition $S_\psi$.*

*Proof (Outline.). Given any axiom $\psi$ in table 2, it is a known result that $F_S \models \psi$ if and only if $F_S$ has the Kripke property $R_\varphi$ shown in table 2 (see [7] for details). But since validity on a $\mathcal{VSK}$ system $S$ is defined in terms of the generated frame $F_S$, the equivalence between items 1 and 2 follows.*

*For each line of the table, the equivalence between 2 and 3 can be established by re-writing the relational properties on Kripke frames in terms of the $\mathcal{VSK}$ conditions on $\mathcal{VSK}$ systems.*

**Other Interaction Axioms** Before we leave our study of $\mathcal{VSK}$ interaction axioms, it is worth noting that there are many other possible interaction axioms of interest [7]. The most important of these have the following general form.

$$\boxplus_i \varphi \Rightarrow \boxdot_j \boxplus_i \varphi \qquad \text{where } \boxplus_i \in \{\mathcal{S}_i, \mathcal{V}_i, \mathcal{K}_i\}, \boxdot_j \in \{\mathcal{S}_j, \mathcal{V}_j, \mathcal{K}_j\}, i \neq j. \tag{11}$$

It is easy to see that schema (11) generates nine possible interaction axioms. We can prove the following general result about such interaction axioms.

**Lemma 2.** *For any system $S$, we have that the generated frame $F_S$ satisfies the following property.*

$$F_S \models \boxplus_i p \Rightarrow \boxdot_j \boxplus_i p \text{ if and only if } \sim_i^{\boxplus} \subseteq \sim_j^{\boxdot}$$

*where $\boxplus_i \in \{\mathcal{S}_i, \mathcal{V}_i, \mathcal{K}_i\}$, $\boxdot_j \in \{\mathcal{S}_j, \mathcal{V}_j, \mathcal{K}_j\}$ and $\sim_i^{\boxplus}$ (respectively $\sim_j^{\boxdot}$) is the equivalence relation corresponding to the modal operator $\boxplus_i$ (respectively $\boxdot_j$).*

*Proof.* Follows from the results presented in [7, Lemma A.11].

Thanks to the above result we can prove that the classes of $\mathcal{VSK}$ systems analysed above are also characterised by the axioms discussed in this section. Indeed we have the following.

**Corollary 1.** *For any axiom $\psi$ of table 3 and any $\mathcal{VSK}$ system $S$ we have that the following are equivalent:*

1. *The system $S$ validates $\psi$, i.e., $S \models \psi$;*
2. *The generated frame $F_S$ satisfies the corresponding Kripke condition $R_\psi$;*
3. *The system $S$ satisfies the corresponding $\mathcal{VSK}$ condition $S_\psi$.*

*Proof. Follows from Lemma 2 and Theorem 1.*

## 5 A Case Study

In order to illustrate the use of multi-agent $\mathcal{VSK}$ logic, we consider again the scenario presented in section 1. While the scenario can be equally explored by means of $\mathcal{VSK}$ semantics, here we focus on the axiomatic side of the formalism.

As discussed in section 1, we have three robotic agents $A, B, C$ involved in a coordination problem in a navigation scenario. We suppose the autonomous robots $A, B$ to be equipped with sonars that can perfectly perceive the environment, up to a certain distance of, say, 1 metre; so their visibiliy function is not transparent (see Table 1). We further admit that within 1 metre of distance of the object the pairing sonar/environment is perfect; hence within this distance the environment is fully visible. For the ease with which we assume it is possible to process signals from sensors, we further assume that if the sensors are adequately working, then the agents have perfect perception, i.e. they are semantically described by a perfect *see* function as in Table 1. We also assume that agents know everything they see, i.e. that their $\tau$ function is local.

Further assume that the robots $A, B$ follow the following rule: if they know that there is a moving object apparently about to collide with them, then they must take evasive action either when this is the only way to avoid a collision, or in case the object is another robot, when this has right of way. This rules are commonly known, or at least that they hold however nested in a number of $\mathcal{K}$ operators. The superuser has access to the sensors of all the agents (it therefore sees what the agents see and knows what is visible to the agents — see previous section) plus some fixed sensors in the environment they inhabit. Hence we model agent $C$ by supposing that it has perfect perception of the environment, that the environment is completely visible to it and that all its perceptions are known by it.

We can now tailor the specification above to the scenario currently in analysis. We have that agents $A, B$ are in a collision course with $A$ having right of way, that this is visible both to agent $B$ and to agent $A$, except that while agent $A$ does see this, agent $B$ does not. Formally:

$$\vdash coll \wedge \mathcal{V}_A coll \wedge \mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll \wedge r\text{-}o\text{-}w_A.$$

Given the assumptions on the agents presented above, it is possible to show that it follows that agent $A$ will take evasive action and that agent $B$ will be shut down by the controller agent $C$. A proof of this is as follows:

| | | |
|---|---|---|
| 1. | $\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll \wedge \mathcal{V}_A coll \wedge r\text{-}o\text{-}w_A$ | [Given] |
| 2. | $\mathcal{V}_C(\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll) \Rightarrow \mathcal{S}_C(\mathcal{V}_B coll \wedge$ $\neg \mathcal{S}_B coll) \Rightarrow \mathcal{K}_C(\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll)$ | [Perfect Perception] |
| 3. | $\mathcal{K}_C(\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll) \Rightarrow shutdown_B$ | [Given] |
| 4. | $(\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll) \Rightarrow \mathcal{K}_C(\mathcal{V}_B coll \wedge \neg \mathcal{S}_B coll)$ | [Given] |
| 5. | $shutdown_B$ | [1,3,4 + Taut] |
| 6. | $\mathcal{K}_A((\neg ev\text{-}act_B \wedge r\text{-}o\text{-}w_A) \Rightarrow \neg \mathcal{K}_B coll)$ | [Given + Taut] |
| 7. | $\neg ev\text{-}act_B \Rightarrow \mathcal{S}_A \neg ev\text{-}act_B \Rightarrow \mathcal{K}_A \neg ev\text{-}act_B$ | [Perfect Perception] |
| 8. | $\mathcal{K}_A \neg \mathcal{K}_B coll$ | [6, 7, K] |
| 9. | $\mathcal{K}_A(coll \wedge \neg \mathcal{K}_B coll) \Rightarrow ev\text{-}act_A$ | [Given] |
| 10. | $\mathcal{V}_A coll \Rightarrow \mathcal{S}_A coll \Rightarrow \mathcal{K}_A coll$ | [Perfect Perception] |
| 11. | $\mathcal{K}_A r\text{-}o\text{-}w_A$ | [1, Perfect Perception] |
| 12. | $ev\text{-}act_A$ | [1, 8, 9, 10, 11, K] |

## 6 Conclusions

In order to design or understand the behaviour of many multi-agent systems, it is necessary to reason about the *information properties* of the system — what information the agents within it have access to, what they actually perceive, and what they know. In this paper, we have presented a logic for reasoning about such properties, demonstrated the relationship of this logic to an abstract general model of multi-agent systems, and investigated various interaction axioms of the logic. Many issues suggest themselves as candidates for future work: chief among them is completeness. In [8], we proved completeness for a mono-modal fragment of $\mathcal{VSK}$ logic. In particular, we proved completeness not simply with respect to an abstract class of Kripke frames, but with respect to the class of Kripke frames corresponding to our model of agents and environments. It is reasonable to expect the proof to transfer to multi-agent settings. However, when interaction axioms of the form studied in section 4 are present, matters naturally become more complicated, and an analysis for each different system is required. This is future work, as are such issues as temporal extensions to the logic, and complexity results.

## References


1. 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.
2. R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. The MIT Press: Cambridge, MA, 1995.
3. R. Goldblatt. *Logics of Time and Computation, Second Edition, Revised and Expanded*, volume 7 of *CSLI Lecture Notes*. CSLI, Stanford, 1992. Distributed by University of Chicago Press.
4. J. Hintikka. *Knowledge and Belief*. Cornell University Press: Ithaca, NY, 1962.
5. G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, New York, 1996.
6. L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.
7. A. Lomuscio. *Knowledge Sharing among Ideal Agents*. PhD thesis, School of Computer Science, University of Birmingham, Birmingham, UK, June 1999.
8. M. Wooldridge and A. Lomuscio. A logic of visibility, perception, and knowledge: Completeness and correspondence results. In *Formal and Applied Practical Reasoning – Proceedings of the Third International Conference (FAPR-2000)*, 2000.
9. Michael Wooldridge and Alessio Lomuscio. Reasoning about visibility, perception and knowledge. 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.