ALESSIO LOMUSCIO
MAREK SERGOT

Deontic Interpreted Systems

Abstract. We investigate an extension of the formalism of interpreted systems by Halpern and colleagues to model the correct behaviour of agents. The semantical model allows for the representation and reasoning about states of correct and incorrect functioning behaviour of the agents, and of the system as a whole. We axiomatise this semantic class by mapping it into a suitable class of Kripke models. The resulting logic, $\mathrm{KD45}_n^{i-j}$, is a stronger version of KD, the system often referred to as Standard Deontic Logic. We extend this formal framework to include the standard epistemic notions defined on interpreted systems, and introduce a new doubly-indexed operator representing the knowledge that an agent would have if it operates under the assumption that a group of agents is functioning correctly. We discuss these issues both theoretically and in terms of applications, and present further directions of work.

1. Introduction

The design of complex multi-agent systems is increasingly having to confront the possibility that agents may not behave as they are supposed to. In e-commerce, in security, in automatic negotiation, in any application where agents are programmed by different parties with competing interests, it is unrealistic to assume that all agents will behave according to some given protocol or standard of behaviour. In addition to analysing the properties that hold if protocols are followed correctly, it is also necessary to predict, test, and verify the properties that would hold if these protocols were to be violated. It is also necessary to test the effectiveness of introducing proposed control and enforcement mechanisms. For these purposes it is often useful to view the behaviour of agents as governed by *norms*, and to seek to apply the formal tools of deontic logic (the logic of obligation and permission) to represent and reason about the distinction between *ideal* (correct, acceptable) behaviour and *actual* (possibly incorrect, unacceptable) behaviour [JS93].

Formal methods and logic in particular have a long tradition in artificial intelligence and distributed computing. Their role, it is argued, is to provide a precise and unambiguous formal tool to *specify* and *reason about* complex systems. However, they have often been attacked by software engineers because of the allegedly somewhat unclear contribution they make towards the engineering of complex computing systems. One of the criticisms most often aired is that logic specifications do not provide *constructive methodologies* for building distributed systems, and so they can be of only limited significance in practice. These different views have led the fields of theoretical and practical distributed computing to diverge. This situation has not changed in the advent of the Multi-Agent system (MAS) paradigm.

One of the suggestions that have been put forward [Woo00] to make MAS theories more relevant to practitioners is the shift to a semantics which is *computationally grounded*. This remark applies to distributed artificial intelligence in general but it is particularly relevant for the case of MAS theories, where semantics are usually given by borrowing ideas developed originally in analytical philosophy. Indeed, most of the more highly respected theories for modelling knowledge, beliefs, intentions, obligations, communications, etc, in MAS are based upon works done in the second half of the last century in philosophical logic. While cross-fertilisation of fruitful ideas can only be regarded positively, one should note that the semantics developed in philosophical logic, even if appropriate for the original task (which is already a matter of debate), may not be the best option for distributed computing applications.

As is widely known, the semantics commonly used for MAS theories is based on Kripke models [Kri59]. A Kripke model $M = (W, R_1, \ldots, R_n, \pi)$ is a tuple composed of a set W, n relations $R_i \subseteq W \times W$, together with an interpretation π for the atoms of the language. The points W represent possible alternatives of the world, and depending on the application under consideration stand for temporal snapshots of the evolution of the world (temporal logic), epistemic alternatives (epistemic logic), 'ideal' or 'deontically perfect' alternatives (standard deontic logic), and so on. Various modal operators can be interpreted by using this semantics, and a heritage of techniques has been developed to prove meta-logical properties about the logics [BdRV01]. Notwithstanding this, no clear correspondence can be drawn between a Kripke model and a distributed computing system.

It has been argued that this lack of correspondence is a serious drawback for attempts to close the gap between theory and practice. Indeed, without a relevant semantics well studied meta-logical properties such as completeness do not seem to be of relevance, and the only possible point of contact between the theorist and the practitioner seems to be logical formulas representing specifications that the theorist would propose to the practitioner. Appropriate grounded semantics aim at bridging this gap by providing practitioners and theoreticians with a convenient and intuitive semantical tool. A grounded semantics should aim at ensuring that a clear correspondence can, at least in principle, be found between states in the computing system and configurations in the semantical description.

The idea of moving away from Kripke models, while still benefiting from most of its technical apparatus, is not new. Indeed, part of the knowledge representation literature uses modal languages defined on semantic structures called *Interpreted systems* (see Chapter 4 of [FHMV95] for details). The idea is to describe a distributed computing system by specifying the states in which every agent and the environment can find itself. In this setting, the level of abstraction at which one chooses to operate is left open; one has the possibility of adopting a fine grain of detail by describing precisely the protocol that the MAS operates, or one can limit oneself to describing macroscopic properties of the MAS, such as epistemic and temporal properties. If needed, logics similar or equivalent to the ones used in philosophical logic can be defined on these semantics. One obvious advantage is the possibility of studying the logic characterisation of systems defined semantically, as opposed to isolating complete semantical structures for a specification.

In this paper we run the following exercise. We consider the basic notion of interpreted system as defined by Halpern et al. in [FHMV95] and show how it can be trivially adapted to provide a basic grounded formalism for some deontic issues. In particular we aim at representing local and global states of violation and compliance (with respect to some functioning protocol). By using these concepts we would like to give a grounded semantics to the deontic notions of *ideal functioning behaviour* of an agent, or of a system of agents, to the concept of the *knowledge that an agent is permitted to have* (again with respect to an ideal functioning protocol), and to the knowledge that an agent has *on the assumption that components of the system are functioning correctly according to their protocols*. Once this task is accomplished, we axiomatise the resulting semantical class, and discuss some of the properties of the logic it determines.

The rest of the paper is organised as follows. In the next section we fix the notation and point to some basic modal logic facts that will become useful later on. In Section 2, we define deontic interpreted systems, and define satisfaction, and validity, of a modal language on them. In Section 3 we study their axiomatisation. We analyse the theorems and logical properties of the system in Section 4, where we also comment on other related notions. In Section /refapplications we show how the framework can be applied to the analysis of the a widely discussed example in distributed computing, the bit transmission problem. We conclude in Section 7.

2. Deontic interpreted systems

2.1. Syntax

We assume a set P of propositional atoms, and a set $A = 1, \ldots, n$ of agents.

DEFINITION 1. *The language $\mathcal{L}$ is defined as follows.*

$$\varphi ::= \textbf{false} \mid \textit{any element of } P \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{O}_i\, \varphi \quad (i \in A).$$

We use the indexed modal operator $\mathcal{O}_i$ to represent the *correctly functioning circumstances of agent i*: the formula $\mathcal{O}_i\, \varphi$ stands for "in all the possible correctly functioning alternatives of agent i, φ is the case", or "whenever agent i is functioning correctly (with respect to some protocol or specification) φ is the case". The formula φ can either refer to local or global properties or to both at the same time. We write $\mathcal{P}_i$ for the dual of $\mathcal{O}_i$: $\mathcal{P}_i\,\varphi =_{def} \neg\,\mathcal{O}_i\,\neg\varphi$. $\mathcal{P}_i\,\varphi$ can be read as "there is a state where agent i is functioning correctly, and in which φ holds.".

We have chosen the symbol $\mathcal{O}_i$ because its semantics will be similar to that of the obligation operator of standard deontic logic. However, it would not be appropriate to read $\mathcal{O}_i\,\varphi$ as "it is obligatory for agent i that φ". The concept we explore is clearly related to what is discussed in mainstream Deontic Logic, but it is not our aim in this paper to provide a characterisation of the multi-faceted concepts of obligation and permission.

Note. In line with much of the literature, we denote a normal modal logic by listing the axioms that define it, under the assumption that uniform substitution, necessitation, and modus ponens hold. For example by $\mathrm{KT45}_n$ we denote the logic obtained by considering axioms K, T, 4, and 5 for n agents (see [FHMV95] for more details).

2.2. Kripke frames

In the following we assume familiarity with basic modal logic techniques and results. We refer the reader to [HC96, Gol92, BdRV01] for more details. In particular, we record here the following for convenience.

DEFINITION 2. *A frame $F = (W, R_1, \ldots, R_n)$ is* serial *if for any relation R_i we have that for any $w \in W$ there exists a $w' \in W$ such that $w\,R_i\,w'$. A frame $F = (W, R_1, \ldots, R_n)$ is* Euclidean *if for any relation R_i we have that for all $w, w', w'' \in W$, $w\,R_i\,w', w\,R_i\,w''$ implies $w'\,R_i\,w''$. A frame $F = (W, R_1, \ldots, R_n)$ is* transitive *if for any relation R_i we have that for all $w, w', w'' \in W$, $w\,R_i\,w', w'\,R_i\,w''$ implies $w\,R_i\,w''$.*

OBSERVATION 1. *The logic $KD45_n$ is sound and complete with respect to serial, transitive and Euclidean frames.*

DEFINITION 3 (p-morphism). *A* frame p-morphism *from $F = (W, R_1, \ldots, R_n)$ to $F' = (W', R'_1, \ldots, R'_n)$ is a function $p : W \to W'$ such that:*

1. *the function p is surjective,*
2. *for all $u, v \in W$ and each $i = 1 \ldots n$, if $u\,R_i\,v$ then $p(u)\,R'_i\,p(v)$,*
3. *for each $i = 1 \ldots n$ and $u \in W$ and $v' \in W'$, if $p(u)\,R'_i\,v'$ then there exists $v \in W$ such that $u\,R_i\,v$ and $p(v) = v'$.*

If there is a p-morphism from F to F', F' is also said to be a p-morphic image of F.

The following result (see for example [Gol92] for the mono-modal case) shows that p-morphisms preserve satisfaction and validity for the language $\mathcal{L}$.

OBSERVATION 2. *If p is a frame p-morphism from F to F' then for all $\varphi \in \mathcal{L}$, if $F \models \varphi$ then $F' \models \varphi$.*

2.3. Deontic interpreted systems

Interpreted systems were originally defined by Halpern and Moses [HM90], and their potential later investigated in greater detail in [FHMV95]. They provide a general framework for reasoning about properties of distributed systems, such as synchrony, a-synchrony, communication, failure properties of communication channels, etc. One of the reasons for the success of interpreted systems is the ease with which states of knowledge can be ascribed to the agents in the system.

The fundamental notion on which interpreted systems are defined is the one of 'local state'. Intuitively, the local state of an agent represents the entire information about the system that the agent has at its disposal. This may be as varied as to include program counters, variables, facts of a knowledge base, or indeed a history of these. The (instantaneous) state of the system is defined by taking the local states of each agent in the system, together with the local state for the environment. The latter is used to represent information which cannot be coded in the agents' local states such as messages in transit, etc.

More formally, consider n non-empty sets $L_1, \dots, L_n$ of local states, one for every agent of the system, and a set of states for the environment L_e. Elements of L_i will be denoted by $l_1, l'_1, l_2, l'_2, \dots$. Elements of L_e will be denoted by $l_e, l'_e, \dots$.

DEFINITION 4 (System of global states). *A* system of global states for n agents S *is a non-empty subset*[1] *of the Cartesian product $L_e \times L_1 \times \cdots \times L_n$.*

An interpreted *system of global states is a pair (S, π) where S is a system of global states and $\pi : S \to 2^P$ is an interpretation function for the atoms.*

The framework presented in [FHMV95] represents the temporal evolution of a system by means of *runs*; these are functions from the natural numbers to the set of global states. An *interpreted system*, in their terminology, is a set of runs over global states together with a valuation for the atoms of the language on points of these runs. In this paper we do not deal with time, and so we will simplify this notion by not considering runs.

[1]The case of the full Cartesian product was analysed in [LMR00].

We now define *deontic systems of global states* by assuming that for every agent, its set of local states can be divided into allowed and disallowed states. We indicate these as *green states*, and *red states* respectively.

DEFINITION 5 (Deontic system of global states). *Given n agents and $n+1$ mutually disjoint and non-empty sets $G_e, G_1, \ldots, G_n$, a* deontic system of global states *is any system of global states defined on $L_e \supseteq G_e, \ldots, L_n \supseteq G_n$. G_e is called the* set of green states for the environment*, and for any agent i, G_i is called* the set of green states for agent i*. The complement of G_e with respect to L_e (respectively G_i with respect to L_i) is called the* set of red states for the environment (respectively for agent i*).*

Given an agent, red and green local states respectively represent 'disallowed' and 'allowed' states of computation. An agent is in a disallowed state if this is in contravention of its specification, as is the case, for example, in a local system crash, or a memory violation. The notion is quite general however: classifying a state as 'disallowed' (red) could simply signify that it fails to satisfy some desirable property. In applications to specific examples it is often useful to classify as red the states that result from the failure of an agent to follow its functioning protocol. In these cases one can consider a finer-grained notion of interpreted systems in which the concepts of protocols and transitions are introduced. Moreover, a rather different and interesting approach is to label runs of the system as 'red' or 'green' instead of states, enabling us to reason about allowed/acceptable as opposed to disallowed/unacceptable/faulty runs. We do not consider these further issues in this paper.

Note that any collection of red and green states as above identifies a *class* of global states. The class of deontic systems of global states is denoted by $\mathcal{DS}$.

DEFINITION 6 (Interpreted deontic system of global states). *An* interpreted deontic system of global states *IDS for n agents is a pair $IDS = (DS, \pi)$, where DS is a deontic system of global states, and π is an interpretation for the atoms.*

In the knowledge representation literature interpreted systems are used to ascribe knowledge to agents, by considering two global states to be indistinguishable for an agent if its local states are the same in the two global states. Effectively, this corresponds to generating a Kripke frame from a system of global states (some formal aspects of this mapping have been explored in [LR98]). In this case, the relations on the generated Kripke frame are equivalence relations; hence (see [FHMV95]) the logic resulting by defining a family of modal operators representing a 'bird's eye view' of the knowledge of the agents is $\mathrm{S5}_n$.

In this paper we set out to do a similar exercise. We investigate how to axiomatise deontic systems of global states using the languages defined in Definition 1,

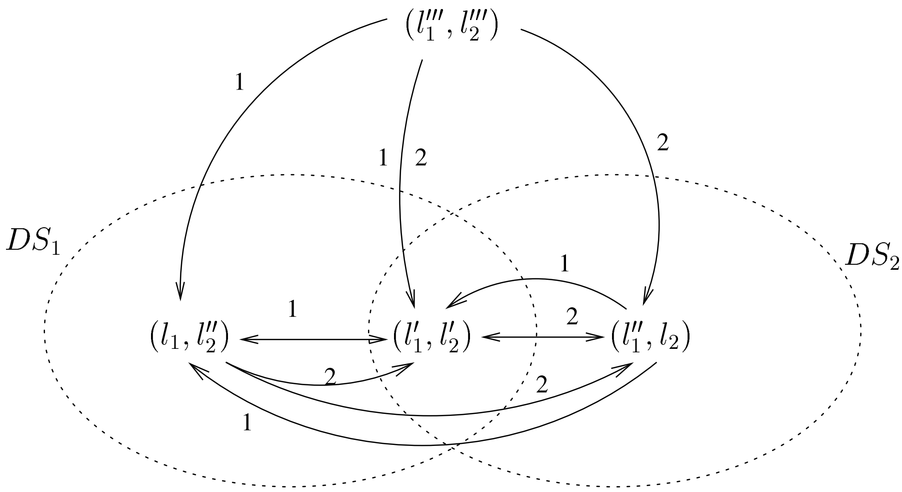

Figure 1. An example of deontic system and its generated frame. In the example above the environment is not considered and the local states for the agents are composed as follows. Agent 1: $L_1 = \{l_1, l_1', l_1'', l_1'''\}, G_1 = \{l_1, l_1'\}$. Agent 2: $L_2 = \{l_2, l_2', l_2'', l_2'''\}, G_2 = \{l_2, l_2'\}$. $DS = \{(l_1', l_2'), (l_1, l_2''), (l_1'', l_2), (l_1''', l_2''')\}$. In the figure the sets DS_1, DS_2 represent the subsets of DS which present acceptable configurations respectively for agent 1, and 2. The labelled links indicate the relations R_1 and R_2 of the generated frame.

and study the properties of the resulting formalisation. In the spirit of the interpreted systems literature we interpret modal formulas on the Kripke models that are built from deontic systems of global states. In order to do this, we first define the frame generated by a deontic system.

DEFINITION 7 (Frame generated by a system). *Given a deontic system of global states DS, the* generated frame $F(DS) = (W, R_1, \ldots, R_n)$ *is defined as follows.*

- $W = DS$.
- *For any $i = 1, \ldots, n$, $\langle l_e, l_1, \ldots, l_n \rangle \, R_i \, \langle l_e', l_1', \ldots, l_n' \rangle$ if $l_i' \in G_i$.*

The function F is naturally extended to map interpreted systems of global states to Kripke models as follows: if $F(DS) = (W, R_1, \ldots, R_n)$ then $F(DS, \pi) = (W, R_1, \ldots, R_n, \pi)$.

Intuitively, the relations R_i represent an accessibility function to global states in which agent i is running according 'correct (or acceptable) operating circumstances'. We illustrate this in Figure 1.

We make use of the construction above to give an interpretation to the formulas of a language as follows. Given an interpreted deontic system $IDS = (DS, \pi)$, the interpretation of formulas of the language $\mathcal{L}$ is defined on the corresponding generated Kripke model $F(DS, \pi)$, where the truth of a formula $\mathcal{O}_i\, \varphi$ at a global state signifies the truth of formula φ at all i-related worlds, i.e., at all the global states in which i is in a correct ('green') local state.

DEFINITION 8 (Satisfaction on interpreted deontic systems of global states). *For any $\varphi \in \mathcal{L}$, $g \in DS$, and $IDS = (DS, \pi)$, satisfaction is defined by:*

$$IDS \models_g \varphi \text{ if } F(DS, \pi) \models_g \varphi,$$

where this is defined as:

$$\begin{array}{ll} F(DS,\pi) \models_g \textbf{true} & \\ F(DS,\pi) \models_g p & \text{if } g \in \pi(p) \\ F(DS,\pi) \models_g \neg\varphi & \text{if not } F(DS,\pi) \models_g \varphi \\ F(DS,\pi) \models_g \mathcal{O}_i\, \varphi & \text{if for all } g' \text{ we have that } g\, R_i\, g' \text{ implies } F(DS,\pi) \models_{g'} \varphi. \end{array}$$

In other words, the truth of formula $\mathcal{O}_i\, \varphi$ at a global state signifies the truth of formula φ in all the global states in which agent i is in a correct local state, i.e. in a green state.

Validity on deontic systems is defined similarly.

DEFINITION 9 (Validity on deontic systems). *For any $\varphi \in \mathcal{L}$, and $IDS = (DS, \pi)$, validity on interpreted deontic systems of global states is defined by $IDS \models \varphi$ if $F(DS, \pi) \models \varphi$. For any $\varphi \in \mathcal{L}$, and $DS \in \mathcal{DS}$, validity on deontic systems of global states is defined by $DS \models \varphi$ if $F(DS) \models \varphi$.*

For any $\varphi \in \mathcal{L}$, we say that φ is valid *on the class $\mathcal{DS}$, and write $\mathcal{DS} \models \varphi$, if for every $DS \in \mathcal{DS}$ we have that $DS \models \varphi$.*

In the following we investigate the logical properties that deontic systems of global states inherit. From Definition 9 it follows that this analysis can be carried out on the class of the generated frames.

3. Axiomatisation

In this section we study deontic systems of global states from the axiomatic point of view. An immediate consideration comes from the following.

LEMMA 1. *Given any DS, we have that $F(DS)$ is serial, transitive, and Euclidean.*

PROOF. $F(DS)$ is serial: this follows from the assumption that for any $i \in A$, we have that $G_i \neq \emptyset$.

$F(DS)$ is transitive: assume $g\, R_i\, g'$, and $g'\, R_i\, g''$, for some $i \in A$. But then, by definition, it must be that $g' = \langle l'_e, \ldots, l'_i, \ldots, l'_n \rangle$, and $l'_i \in G_i$. Similarly, it must be that $g'' = \langle l''_e, \ldots, l''_i, \ldots, l''_n \rangle$, and $l''_i \in G_i$. But then it must be that $g\, R_i\, g''$.

$F(DS)$ is Euclidean: assume $g\, R_i\, g'$, and $g\, R_i\, g''$, for some $i \in A$. So, it must be that $g'' = \langle l''_e, \ldots, l''_i, \ldots, l''_n \rangle$, and $l''_i \in G_i$. So, we have $g'\, R_i\, g''$. ■

This observation leads immediately to the conclusion that the logic of deontic systems of global states must be at least as strong as KD45$_n$, which is to be expected. However, as will be clearer in the following, it turns out that the logic determined by deontic systems of global states is in fact stronger than KD45$_n$. In order to see this, we need to introduce a few semantic structures.

3.1. Some secondary properties of Kripke frames

In order to obtain an axiomatisation for deontic systems of global states, we introduce some *secondary properties* of Kripke frames, by which we mean properties that hold on any sub-frame that can be reached from some point in the frame. (The term 'secondarily reflexive' is used in [Che80, p92].)

Notation For a binary relation R on W and $w \in W$, $R(w)$ denotes the set of points in W that are R-accessible from w, i.e., $R(w) =_{def} \{w' \in W \mid w\, R\, w'\}$.

LEMMA 2. *Let R be a binary relation on W. R is Euclidean iff R is universal on $R(w)$ for all $w \in W$.*

PROOF. Suppose $w_1 \in R(w)$ and $w_2 \in R(w)$. Then (by definition) $wR\, w_1$ and $wR\, w_2$. But then $w_1 R\, w_2$ (R is Euclidean). For the other half: suppose $wR\, w_1$ and $wR\, w_2$. Then $w_1 \in R(w)$ and $w_2 \in R(w)$, and so $w_1 R\, w_2$ since R is universal on $R(w)$. ■

DEFINITION 10 (Secondarily universal). *Let R be a binary relation on W. R is* secondarily universal *if*

(i) for all $w \in W$, R is universal on $R(w)$;

(ii) for all $w', w'' \in W$, $R(w') = R(w'')$.

A frame $F = (W, R_1, \ldots, R_n)$ is a secondarily universal frame if every relation R_i, $i \in A$, is secondarily universal.

It follows (by Lemma 2) that condition (i) of the definition is equivalent to the requirement that R is Euclidean. We have then that every secondarily universal relation is Euclidean.

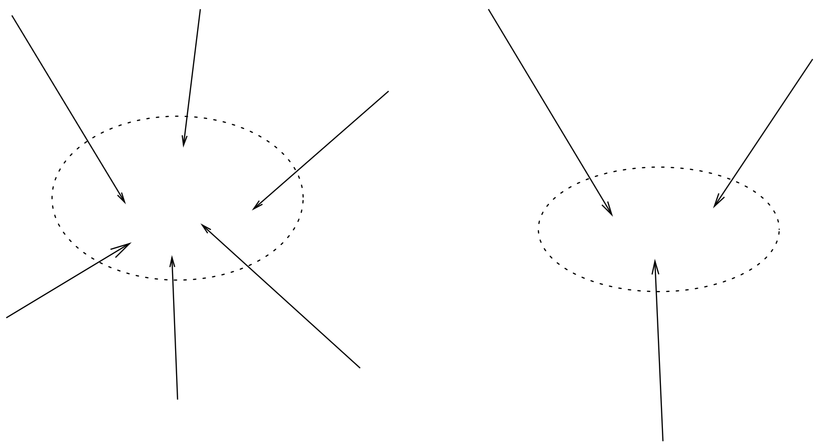

Figure 2. An example of a secondary universal frame. The arrows represent relations between points (not shown), and the dotted ellipses represent sets of points in which each point is related to all the points in the dotted ellipse. Note that a secondary universal frame may contain a number of these unconnected sub-frames.

Example Consider $R = \{(w_1, w_2), (w_2, w_2), (w_3, w_3), (w_2, w_3), (w_3, w_2)\}$. This R is not secondarily universal. $R(w_1) = \{w_2\}$, $R(w_2) = R(w_3) = \{w_2, w_3\}$. $R(w_1) \neq R(w_2)$. $R' = R \cup \{(w_1, w_3)\}$ is secondarily universal (assuming R' is a relation on the set $\{w_1, w_2, w_3\}$). The relation R'' on the set $\{w_1, w_2, w_3, w_4\}$ where $R'' = R' \cup \{(w_4, w_2), (w_4, w_3)\}$ is also secondarily universal.

LEMMA 3. *Let R be a binary relation on W. If R is secondarily universal then R is transitive.*

PROOF. Suppose wRw' and $w'Rw''$. Then $w' \in R(w)$ and $w'' \in R(w')$. But $R(w) = R(w')$, so $w'' \in R(w)$, i.e., wRw''. ∎

Now we have a very useful representation result.

THEOREM 1 (Secondarily universal: Representation theorem). *Let R be a binary relation on W. R is secondarily universal if and only if there exists a subset $S \subseteq W$ such that, for all $w, w' \in W$, wRw' iff $w' \in S$.*

Furthermore, R is serial if and only if S is non-empty.

PROOF. Right to left is easy to check and omitted here.

Left to right: If R is empty, take $S = \emptyset$. The result holds trivially. If R is not empty, take $S = R(\overline{w})$ for some $\overline{w} \in W$. Consider any $w, w' \in W$. wRw' iff $w' \in R(w)$ (by definition). But R secondarily universal implies $R(w) = R(\overline{w}) = S$. So wRw' iff $w' \in S$ as required. And clearly if R is non-empty then S is non-empty; from its definition, it also then follows that R is serial. ∎

We are now in a position to relate validity on the class of serial secondarily universal frames to validity on the class of serial, transitive and Euclidean frames. However, we are interested here in the multi-modal case, and for this we need a property of frames we call *i-j Euclidean*.

DEFINITION 11 (i-j Euclidean frame). *A frame* $F = (W, R_1, \ldots, R_n)$ *is* i-j Euclidean *iff for all* $w, w', w'' \in W$, *and for all* $i, j \in A$, *we have that* $w \, R_i \, w', w \, R_j \, w''$ *implies* $w'' \, R_i \, w'$.

The class of i-j Euclidean frames collapses to 'standard' Euclidean frames for $i = j$.

There is a precise correspondence that can be drawn between i-j Euclidean frames and the following axiom:

$$\mathcal{P}_i \, p \to \mathcal{O}_j \, \mathcal{P}_i \, p \qquad \text{(for any } i, j \in A) \qquad 5^{i\text{-}j}$$

LEMMA 4. *A frame* F *is i-j Euclidean if and only if* $F \models 5^{i\text{-}j}$.

PROOF. From left to right. Consider any i-j Euclidean model $M = (F, \pi)$, such that $M \models_w \mathcal{P}_i \, p$, so there exists a point w' such that wR_iw', and $M \models_{w'} p$. It remains to prove that for any w'' such that wR_jw'' it must be that $M \models_{w''} \mathcal{P}_i \, p$. But F is i-j Euclidean, so we have $w''R_iw'$, which is what we needed.

From right to left. Consider three points $w, w', w'' \in W$ such that wR_iw', wR_jw''. Consider a valuation π such that $\pi(p) = \{w'\}$, and the model $M = (F, \pi)$. we have $M \models_w \mathcal{P}_i \, p$. We have that $M \models_w \mathcal{O}_j \, \mathcal{P}_i \, p$; therefore it must be that $M \models_{w'} \mathcal{P}_i \, p$; but then it must be that $w''R_iw'$. ■

Now we will relate validity on the class of (serial) secondarily universal frames to validity on the class of (serial) transitive, i-j Euclidean frames.

LEMMA 5. *If a frame* F *is secondarily universal then it is also i-j Euclidean.*

PROOF. Consider any relations R_i and R_j of the frame F. R_i and R_j are secondarily universal, and so by the representation Theorem 1, there exist subsets S_i and S_j of W such that, for all $w, w', w'' \in W$, wR_iw' iff $w' \in S_i$ and wR_jw'' iff $w'' \in S_j$. So now: wR_iw' and wR_jw'' implies $w'' \in S_j$, and hence $w'R_jw''$. ■

Notation For frame $F = (W, R_1, \ldots, R_n)$ we write wR^*w' when w' can be reached from w by a path of finite length (including zero) using any combination of relations $R_1, \ldots, R_n$, i.e. (more precisely) when R^* is the reflexive transitive closure of $\cup_{i=1}^{n} R_i$.

With this notation, the standard notion of a generated (sub-)frame is expressed as follows.

DEFINITION 12 (Generated frame). *Let w be any world in a frame $F = (W, R_1, \ldots, R_n)$. Then $F^w = (W^w, R_1^w, \ldots, R_n^w)$ is the* frame generated by w from F *when*

- *$W^w = R^*(w)$ (i.e., W^w is the set of worlds accessible from w by any path of finite length of relations $R_1, \ldots, R_n$);*
- *each R_i^w is the restriction of R_i to W^w, i.e., $R_i^w = R_i \cap (W^w \times W^w)$.*

Let $\mathcal{F}$ be any class of frames. $\mathcal{G}(\mathcal{F})$ is the class of frames generated by any point of any frame in $\mathcal{F}$. $\mathcal{F} = \{F^w \mid F \in \mathcal{F}, w \in W\}$.[2]

Now we have the standard result:

OBSERVATION 3. *Let $\mathcal{F}$ be any class of frames and let $\mathcal{G}(\mathcal{F})$ be the class of generated frames. Then for all $\varphi \in \mathcal{L}$, φ is valid on the class of frames $\mathcal{F}$ if and only if φ is valid on the class of frames $\mathcal{G}(\mathcal{F})$.*

PROOF. Straightforward, by induction on the structure on φ. ■

COROLLARY 4. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be classes of frames such that $\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{G}(\mathcal{F}_1)$. Then for all $\varphi \in \mathcal{L}$, $\mathcal{F}_1 \models \varphi$ if and only if $\mathcal{F}_2 \models \varphi$.*

PROOF. From $\mathcal{F}_1 \supseteq \mathcal{F}_2$ it follows by definition that $\mathcal{F}_1 \models \varphi$ implies $\mathcal{F}_2 \models \varphi$. From $\mathcal{F}_2 \supseteq \mathcal{G}(\mathcal{F}_1)$ it follows that $\mathcal{F}_2 \models \varphi$ implies $\mathcal{G}(\mathcal{F}_1) \models \varphi$. But $\mathcal{G}(\mathcal{F}_1) \models \varphi$ implies $\mathcal{F}_1 \models \varphi$. So we have $\mathcal{F}_1 \models \varphi$ implies $\mathcal{F}_2 \models \varphi$ which in turn implies $\mathcal{F}_1 \models \varphi$. ■

We will make use of this corollary in the proof of the main theorem of this section. A couple of further lemmas are also required. We report them in the Appendix, and refer to them in the proofs below when required.

We now prove that the class of serial, transitive, i-j Euclidean frames and the class of serial, secondarily universal frames are semantically equivalent, that is, the same set of formulas φ is valid on both.

THEOREM 2. *For all $\varphi \in \mathcal{L}$, φ is valid on the class of serial, secondarily universal frames if and only if φ is valid on the class of serial, transitive and i-j Euclidean frames.*

[2] We are using the term "generated frame" to denote both a frame generated *by a system* (Definition 7), and a frame generated by a world from a frame (Definition 12). We trust any ambiguity is resolved by the context.

PROOF. Since every secondarily universal frame is also transitive and i-j Euclidean (Lemmas 3 and 5), it suffices by Corollary 4 to prove that the class of frames generated from serial, transitive and i-j Euclidean frames is contained in the class of serial, secondarily universal frames, that is, that every frame that is generated by some world from a serial, transitive and i-j Euclidean frame is also secondarily universal.

Let $F^{\overline{w}}$ be a frame generated from some world $\overline{w}$ of a serial, transitive and i-j Euclidean frame F. We show that $F^{\overline{w}}$ is serial and secondarily universal by showing that, for each $i \in A$, $wR_i^{\overline{w}}w'$ iff $w' \in R_i^{\overline{w}}(\overline{w})$ and $R_i^{\overline{w}}(\overline{w})$ is non-empty; the result follows by the representation Theorem 1. Since F is serial, every R_i is serial and hence $R_i^{\overline{w}}(\overline{w})$ is non-empty. It remains to show that for all $w, w' \in R^*(\overline{w})$, wR_iw' iff $w' \in R_i(\overline{w})$. For the first half: $\overline{w}R^*w$ and wR_iw' implies $\overline{w}R_iw'$ by Lemma 6, and hence that $w' \in R_i(\overline{w})$. For the other half, suppose $\overline{w}R^*w$ and $w' \in R_i(\overline{w})$. Then $\overline{w}R^*w$ and $\overline{w}R_iw'$, which implies wR_iw' by Lemma 7. ■

THEOREM 3. *The logic KD45$_n^{i\text{-}j}$ is sound and complete with respect to*

- *serial, transitive and i-j Euclidean frames*
- *serial, secondarily universal frames.*

PROOF. We show the first part; the second part follows immediately by Theorem 2. We show that the logic KD45$_n^{i\text{-}j}$ is canonical, i.e. that the frame F_C of the canonical model M_C is serial, transitive, and i-j Euclidean. Since the logic in question is stronger than KD45$_n$, from the literature we know that F_C is serial and transitive; we show it is i-j Euclidean. To do so, consider three maximal consistent sets w, w', w'', such that wR_iw', wR_jw''. It remains to show that $w''R_iw'$. By contradiction suppose this is not the case; then there must exist a formula $\alpha \in \mathcal{L}$ such that $\mathcal{O}_i\, \alpha \in w''$, and $\neg\alpha \in w'$. But then $\mathcal{P}_i\, \neg\alpha \in w$, and so $\mathcal{O}_j\, \mathcal{P}_i\, \neg\alpha \in w$, which in turn implies $\mathcal{P}_i\, \neg\alpha \in w''$, i.e. $\neg\, \mathcal{O}_i\, \alpha \in w''$, which is absurd, because it would make w'' inconsistent. ■

Before we can axiomatise deontic systems of global states we need to make clear the correspondence between deontic systems of global states and secondarily universal frames.

THEOREM 4. *Any serial, secondarily universal frame is the p-morphic image of the frame generated by an appropriate deontic system of global states.*

PROOF. Let $F = (W, R_1, \dots, R_n)$ be any serial secondarily universal frame; we define a deontic system DS as follows. Pick some $\overline{w} \in W$. For any relation $i \in A$, let $G_i = R_i(\overline{w})$. Since R_i is serial, this satisfies the requirement that G_i is not empty. Let $G_e = W$.

Consider now the deontic system DS defined as $DS = \{(w, w, \ldots, w) \mid w \in W\}$. For simplicity we use the shortcut $(w, w, \ldots, w) = [w]$. The frame generated by DS is defined (see Definition 7) by $F(DS) = F' = (W', R'_1, \ldots, R'_n)$, where $W' = \{[w] \mid w \in W\}$, and for any $i \in A$ we have $R'_i = \{([w], [w']) \mid w' \in G_i\}$.

We show that F can be seen as the target of a p-morphism of domain F'. Define the function $p : W' \to W$ such that $p([w]) = w$. We prove that p is a p-morphism as defined in Definition 3.

- p is clearly surjective.
- For any $i \in A$, consider $[w]\, R'_i\, [w']$. By definition it must be $w' \in G_i$. So $\forall w'' \in W$ we have that $w''\, R_i\, w'$. But then in particular $w\, R_i\, w'$.
- Consider $w\, R_i\, p([w'])$. So, by construction, we have $p[w'] = w'$, and $w' \in G_i$. But then $[w]\, R'_i\, [w']$.

■

For the result presented in this paper, the notion of p-morphism is enough to achieve the result, but it can be noted that the function defined above is actually an isomorphism.

We can now prove the main result of this section.

THEOREM 5. *The logic KD45$_n^{i\text{-}j}$ is sound and complete with respect to deontic systems of global states.*

PROOF. The proof for soundness is straightforward and omitted here. For completeness, we prove the contrapositive. Suppose $\not\vdash \varphi$; then by Theorem 3, there exists a serial, secondarily universal model $M = (F, \pi)$ such that $M \not\models_w \varphi$, for some $w \in W$. By Theorem 4 there exists a deontic system DS such that $F(DS)$ is the domain of a p-morphism $p : F(DS) \to F$. But then by Lemma 2, since $F \not\models \varphi$, we have that $F(DS) \not\models \varphi$, so $DS \not\models \varphi$, so $\mathcal{DS} \not\models \varphi$, which is what we needed to show. ■

4. Discussion

4.1. The logic KD45$_n^{i\text{-}j}$

In the previous section we showed that the logic KD45$_n^{i\text{-}j}$ provides a complete axiomatisation for deontic systems of global states. In the following we look at the individual axioms in a little more detail.

In light of much of the literature in this area the logic above should be seen as providing a *bird's eye view* of the properties of the MAS. Therefore validity of axiom K:

$$\mathcal{O}_i(p \rightarrow q) \rightarrow (\mathcal{O}_i\, p \rightarrow \mathcal{O}_i\, q) \qquad \text{K}$$

seems reasonable. Indeed, if agent i's functioning specification requires that whenever p is the case then q should also be the case, then, if according to the agent's functioning protocol p is the case, then q should also be the case according to that protocol.

Axiom D guarantees that individual specifications are consistent:

$$\mathcal{O}_i\, p \rightarrow \neg\, \mathcal{O}_i\, \neg p \qquad \text{D}$$

Another way of seeing the above is to note that in normal modal logics, axiom D is equivalent to $\neg\, \mathcal{O}_i$ **false**. Axiom D is sometimes called the characteristic deontic axiom: together with axiom K, axiom D is the basis for Standard Deontic Logic (SDL).

Moving to the next pair of axioms, if we give a bird's eye view reading of the $\mathcal{O}_i$ modality, axiom 4

$$\mathcal{O}_i\, p \rightarrow \mathcal{O}_i\, \mathcal{O}_i\, p \qquad 4$$

and axiom 5

$$\mathcal{P}_i\, p \rightarrow \mathcal{O}_i\, \mathcal{P}_i\, p \qquad 5$$

are perhaps not as strong as a first reading might suggest.

Another way of reading axiom 4 is to note that it forbids the situation in which p is prescribed but it is allowed that p is not prescribed. This seems reasonable with respect to strong deontic notions such as the one we are modelling. For example consider the case of one agent running a program in which one of its variables is supposed to be 'guarded', say to a boolean value. It would then be unreasonable if the protocol were to specify that the variable has to be a boolean, but at the same time allowed it not to be prescribed that it be a boolean. It is worth pointing out that the underlying reason for the validity of axioms 4 and 5 in this context is that the criterion for what counts as a green state is *absolute*, that is to say, the set of green states for an agent is independent of the state in which it currently is. An alternative would be to introduce functions $g_i : L_i \rightarrow 2^{L_i}$ to identify green states; but that seems to have less appeal in the present context and we do not explore it further.

Lastly, axiom $5^{i\text{-}j}$ of the previous section, of which axiom 5 is a special case, also reflects the absolute nature of the specification of 'green'. It represents an interaction between the states of correctly functioning behaviour of pairs of agents.

$$\mathcal{P}_i\, p \rightarrow \mathcal{O}_j\, \mathcal{P}_i\, p \qquad 5^{i\text{-}j}$$

$5^{i\text{-}j}$ expresses the property that if a state of the system can happen under the correct behaviour of one agent i, then the protocol of any agent j must allow this eventuality in any correct state that it specifies for j. Again, this seems a reasonable assumption. Suppose that agent i can follow its functioning protocol and reach a state coded by p. Axiom $5^{i\text{-}j}$ stipulates that in this case agent j's protocol cannot prescribe as admissible any states in which agent i does not have the opportunity to move to a state coded by p. In other words, axiom $5^{i\text{-}j}$ asserts a sort of *independence* in the interplay between agents. Naturally, we do not have the very strong property that all specifications are mutually consistent: $\mathcal{O}_i\, p \rightarrow \neg\, \mathcal{O}_j\, \neg p$ is *not* valid. However, $5^{i\text{-}j}$ provides a weak kind of mutual consistency: agent j's protocol cannot forbid the possibility of p for agent i if this is granted by agent i's protocol.

It is instructive to note that the logic $\mathrm{KD45}_n^{i\text{-}j}$ contains also the following generalisation of axiom 4:

$$\mathcal{O}_i\, p \rightarrow \mathcal{O}_j\, \mathcal{O}_i\, p \qquad 4^{i\text{-}j}$$

This can be checked semantically, or derived as follows: $\mathcal{O}_i\, p \rightarrow \mathcal{O}_i\, \mathcal{O}_i\, p$ (by 4); $\mathcal{O}_i\, \mathcal{O}_i\, p \rightarrow \mathcal{P}_i\, \mathcal{O}_i\, p$ (by D); $\mathcal{P}_i\, \mathcal{O}_i\, p \rightarrow \mathcal{O}_j\, \mathcal{P}_i\, \mathcal{O}_i\, p$ (by $5^{i\text{-}j}$); $\mathcal{O}_j\, \mathcal{P}_i\, \mathcal{O}_i\, p \rightarrow \mathcal{O}_j\, \mathcal{O}_i\, p$ (by the rule RK and the dual of $5^{i\text{-}j}$).

It is now easy to check that the logic $\mathrm{KD45}_n^{i\text{-}j}$ contains all axioms in the following scheme:

$$X_i\, p \leftrightarrow Y_j\, X_i\, p$$

where X_i is any one of $\mathcal{O}_i, \mathcal{P}_i$ and Y_j is any one of $\mathcal{O}_j, \mathcal{P}_j$. There are thus only $2(2n+1)$ distinct modalities in the logic $\mathrm{KD45}_n^{i\text{-}j}$.

4.2. Alternative characterisation: reduction strategies

It is instructive to consider also an alternative characterisation of the logic of deontic systems of global states in a manner analogous to the well-known Andersonian reduction of Standard Deontic Logic to alethic modal logic [And58].

Suppose we augment the language $\mathcal{L}$ of Definition 1 with a modal operator $\Box$ to represent what holds in all global states and a set $\mathbf{g}_1, \ldots, \mathbf{g}_n$ of distinguished propositional constants. Each $\mathbf{g}_i$ is intended to be read as expressing that agent i is in a correctly functioning local state according to its own protocol. We write $\Diamond$ for the dual of $\Box$. The relevant truth conditions are:

$$\begin{array}{ll} F(DS,\pi) \models_g \mathbf{g}_i & \text{if } g \in R_i(g) \quad (i \in A) \\ F(DS,\pi) \models_g \Box\, \varphi & \text{if for all } g', F(DS,\pi) \models_{g'} \varphi. \end{array}$$

The constant $\mathbf{g}_i$ is true in a global state g when agent i is in a correct (green) local state. Expressed directly in terms of the interpreted deontic system $IDS =$

(DS, π), the truth conditions for each $\mathbf{g}_i$ are:

$$(DS, \pi) \models_g \mathbf{g}_i \quad \text{if } l_i(g) \in G_i$$

where l_i is a function that returns i's local state from a global state.

One can see that the truth conditions for $\mathcal{O}_i\, \varphi$ are identical to those for the expression $\Box(\mathbf{g}_i \to \varphi)$. Each operator $\mathcal{O}_i$ can thus be defined as an abbreviation in terms of $\Box$ and $\mathbf{g}_i$ as follows:

$$\mathcal{O}_i\, \varphi =_{def} \Box(\mathbf{g}_i \to \varphi) \qquad \text{Def.}\mathcal{O}_i$$

$\mathcal{P}_i\, \varphi$ is then an abbreviation for $\Diamond(\mathbf{g}_i \wedge \varphi)$.

The model property that every R_i is serial, equivalently that every G_i in the interpreted deontic system is non-empty, validates the following:

$$\neg\Box\neg\mathbf{g}_i \quad \text{i.e.,} \ \Diamond\mathbf{g}_i \qquad \text{D}(\mathbf{g}_i)$$

The logic of $\Box$ is obviously S5 (i.e. type KT5 = KT45). It is easy to check that $\mathcal{O}_i$ as defined above has the properties K, D, 4 ($4^{i\text{-}j}$) and $5^{i\text{-}j}$. For example $4^{i\text{-}j}$ may be derived as follows:

$$\begin{array}{rcll}
\mathcal{O}_i\, p & \to & \Box(\mathbf{g}_i \to p) & (\text{Def.}\mathcal{O}_i) \\
& \to & \Box\Box(\mathbf{g}_i \to p) & (4\Box) \\
& \to & \Box\big(\mathbf{g}_j \to \Box(\mathbf{g}_i \to p)\big) & (\Box \text{ normal}) \\
& \to & \mathcal{O}_j\, \mathcal{O}_i\, p & (\text{Def.}\mathcal{O}_j, \text{Def.}\mathcal{O}_i)
\end{array}$$

The derivation of $5^{i\text{-}j}$ is similar:

$$\begin{array}{rcll}
\mathcal{P}_i\, p & \to & \Diamond(\mathbf{g}_i \wedge p) & (\text{Def.}\mathcal{P}_i) \\
& \to & \Box\Diamond(\mathbf{g}_i \wedge p) & (5\Box) \\
& \to & \Box\big(\mathbf{g}_j \to \Diamond(\mathbf{g}_i \wedge p)\big) & (\Box \text{ normal}) \\
& \to & \mathcal{O}_j\, \mathcal{P}_i\, p & (\text{Def.}\mathcal{O}_j, \text{Def.}\mathcal{P}_i)
\end{array}$$

We also have the following interaction between $\Box$ and each $\mathcal{O}_i$:

$$\Box p \to \mathcal{O}_i\, p \qquad \Box\text{-}\,\mathcal{O}_i$$

Although reduction strategies can be useful to relate and compare the definition of modal operators, note that issues such as completeness cannot be proven through them.

4.3. Notions of 'global' correctness

So far we have described and discussed the use of a green and red state semantics for interpreting the indexed deontic operator of correct behaviour. There are many possible ways to extend these notions to model the notion of *globally correct functioning behaviour* of the MAS. For example, it is straightforward to augment the framework with another modality $\mathcal{O}$ capturing global correctness, interpreted in terms of G, the set of green states for the system as a whole, as follows:

$$F(DS,\pi) \models_g \mathcal{O}\,\varphi \text{ if for all } g' \in G \text{ we have that } F(DS,\pi) \models_{g'} \varphi.$$

How is the notion of global correctness related to local correctness of individual agents or groups of agents? There are many possible relationships, depending on which notion of global correctness we wish to model. For example, the following are the three simplest definitions:

1. $G = \{(l_e, l_1, \ldots, l_n) \mid l_e \in G_e\}$,
2. $G = \{(l_e, l_1, \ldots, l_n) \mid l_i \in G_i \text{ for some } i \in A\}$,
3. $G = \{(l_e, l_1, \ldots, l_n) \mid l_i \in G_i \text{ for all } i \in A\}$,

The first version corresponds to a notion of correct behaviour for the environment. This can be used to model system failures where these are associated with events such as communication breakdown, etc. In the second definition of G, a state of the system is regarded as correct whenever one or more of the agents in the system are in locally correct states; parts of the system might not be performing as intended but parts of it are. This can serve as a guarantee that the system is not completely crashed, as is the case, for example, in a system containing redundant components. It could also perhaps be used to model liveness. The third definition models correct states as states in which all the subcomponents are working correctly. This can be used to model a conservatory notion of correctness, useful when modelling safety critical systems.

Should the second definition from the list above be chosen as semantical model, the resulting axiomatisation would inherit the following interplay between globally and locally correct behaviours:

$$\mathcal{O}\,p \rightarrow \mathcal{O}_i\,p \qquad \text{for some } i \in A.$$

Should the third possibility be adopted, we would inherit the validity of:

$$\mathcal{O}\,p \rightarrow \mathcal{O}_i\,p \qquad \text{for all } i \in A.$$

There are also cases in which a sound notion of global correctness may be unrelated to what the local states are.

It is also straightforward to generalise, to allow for the modelling of arbitrary groups of agents, and not just individual agents and the global system as a whole: $\mathcal{O}_X$ would represent correct functioning of any group of agents $X \subseteq A$, with $\mathcal{O}_X$ interpreted in various ways, in analogous fashion to the different notions of global correctness identified above. The indexed modality $\mathcal{O}_i$ is then the limiting case where X is a singleton $\{i\}$, and global correctness $\mathcal{O}$ is the limiting case where $X = A$. We leave detailed examination of these possibilities to future papers.

5. Deontic states of knowledge

Interpreted deontic systems are an extension of interpreted systems, and as such can be used to interpret knowledge in the same way. To see this, augment the language $\mathcal{L}$ of Definition 1 with an indexed modality K_i representing knowledge of agent i. To give an interpretation to this modality, extend Definition 8 with the following clause:

$$F(DS, \pi) \models_g \mathrm{K}_i\, \varphi \quad \text{if for all } g' \text{ we have that } l_i(g) = l_i(g') \text{ implies } F(DS, \pi) \models_{g'} \varphi,$$

where l_i is a function that returns the i's local state from a global state. It is reasonable to expect that an axiomatisation of the resulting augmented logic will be given by $\mathrm{S5}_n$ for the K_i component union (in the technical sense of [KW91, Gab98, GS98]) the logic $\mathrm{KD45}_n^{i\text{-}j}$ for the deontic part.

What is more interesting though, is that deontic systems of global states allow us to express some more subtle concepts of knowledge not expressible in bare interpreted systems. One of these is the *knowledge that an agent is allowed to have*. Consider, in the first instance, the notion expressed by the construction $\mathcal{O}\,\mathrm{K}_i$. For ease of reference, the truth conditions can be stated equivalently as follows:

$$F(DS, \pi) \models_g \mathcal{O}\,\mathrm{K}_i\, \varphi \quad \text{if for all } g' \in G \text{ we have that } F(DS, \pi) \models_{g'} \mathrm{K}_i\, \varphi.$$

Or:

$$F(DS, \pi) \models_g \mathcal{O}\,\mathrm{K}_i\, \varphi \quad \text{if for all } g', g'' \text{ we have that } l_i(g') = l_i(g'') \text{ and } g'' \in G \text{ implies } F(DS, \pi) \models_{g'} \varphi.$$

Again there are different notions that can be expressed, depending on how we choose to interpret the notion of global correctness modelled by $\mathcal{O}$, that is, what we choose for the specification of the set G of green global states.

It is particularly important when reading these expressions to remember that they express the "bird's eye" view of the MAS: $\mathcal{O}\,\mathrm{K}_i\,\varphi$ says that in all states conforming to correct global behaviour, agent i has sufficient information to know that φ. There are many other notions of 'agent i ought to know φ' that are not captured by this construction. For example, a well-known problem in the study of epistemic obligations is the following observation, sometimes referred to as the paradox of the knower [Åqv67]. Since $\mathrm{K}_i\,\varphi \rightarrow \varphi$ is valid, when $\mathcal{O}$ is normal, we have also validity of the following formula

$$\mathcal{O}\,\mathrm{K}_i\,\varphi \rightarrow \mathcal{O}\,\varphi. \tag{1}$$

So (Åqvist's example) "You ought to know that your wife is committing adultery" implies "it ought to be that Your wife is committing adultery". Clearly, Formula 1 cannot be accepted as a general principle of reasoning about 'ought to know'. An analysis of this problem has driven much of the development of epistemic obligations. With the information theoretical reading of knowledge we use in this paper the formula above is not contentious: it says merely that if in all states conforming to correct global behaviour agent i has sufficient information to know that φ, then in all states conforming to global behaviour φ is the case. Can the account developed in this paper be adapted to give a formalisation of these other senses of 'ought to know'? Perhaps, but this is not something we have explored. Similarly, $\mathcal{O}_j\,\mathrm{K}_i\,\varphi$ expresses that in all states in which agent j is functioning correctly according to its protocol, agent i has the information to know that φ. And likewise for the expression $\mathcal{O}_X\,\mathrm{K}_i\,\varphi$ where X is any subset of the agents A.

A well known paper [GMP92] in the literature of Computer Security by Glasgow et al. attempts to combine deontic and epistemic modalities to express what an agent is permitted to know. The authors provide only a syntactic presentation, making comparisons with this paper hard to draw.

Clearly, we can also study the notions expressed by constructions of the form $\mathrm{K}_i\,\mathcal{O}_j$, $\mathrm{K}_i\,\mathcal{O}$, and $\mathrm{K}_i\,\mathcal{O}_X$. More interesting is a third possibility still, which is knowledge that an agent i has *on the assumption that the system (the environment, agent j, group of agents X) is functioning correctly*. We shall employ the (doubly relativised) modal operator $\widehat{\mathrm{K}}_i^j$ for this notion, interpreted as follows on the interpreted deontic system (DS, π) itself:

$$(DS, \pi) \models_g \widehat{\mathrm{K}}_i^j\,\varphi \quad \text{if for all } g' \text{ such that } l_i(g) = l_i(g') \text{ and } l_j(g') \in G_j \text{ we have that } (DS, \pi) \models_{g'} \varphi,$$

and as follows on the generated frame $F(DS, \pi)$:

$$F(DS, \pi) \models_g \widehat{\mathrm{K}}_i^j\,\varphi \quad \text{if for all } g' \text{ such that } l_i(g) = l_i(g') \text{ and } g' \in R_j(g') \text{ we have that } F(DS, \pi) \models_{g'} \varphi.$$

We write $\widehat{\mathrm{K}}_i$ for the corresponding global analogue: the truth conditions are obtained by replacing the condition $l_j(g') \in G_j$ by $g' \in G$: again, different versions are obtained by choosing among the different options for the definition of what counts as the set of green global states G. And likewise for the obvious generalisation to $\widehat{\mathrm{K}}_i^X$ where X is any (non-empty) subset of the set of agents A.

It is easy to check that the operator $\widehat{\mathrm{K}}_i^j$ satisfies axioms K, 4, and 5 but does not satisfy axiom T. For the epistemic notions modelled by $\widehat{\mathrm{K}}_i^j$, positive and negative introspection (axioms 4 and 5, respectively) do seem reasonable. Moreover, we should not expect that knowledge on the assumption that some other agent or group of agents is behaving correctly implies truth.

It is perhaps clearer to see the relationship between the constructions $\mathcal{O}_j \, \mathrm{K}_i$, $\mathrm{K}_i \, \mathcal{O}_j$ and $\widehat{\mathrm{K}}_i^j$ when they are expressed using the reduction method of section 4.2.

$$\begin{array}{rcl} \mathcal{O}_j \, \mathrm{K}_i \, p & = & \Box(\mathbf{g}_j \rightarrow \mathrm{K}_i \, p) \\ \mathrm{K}_i \, \mathcal{O}_j \, p & = & \mathrm{K}_i \, \Box(\mathbf{g}_j \rightarrow p) \\ \widehat{\mathrm{K}}_i^j \, p & = & \mathrm{K}_i(\mathbf{g}_j \rightarrow p) \end{array}$$

$\mathrm{K}_i \, \mathcal{O}_j$ and $\widehat{\mathrm{K}}_i^j$ are closely related. To see the relationship, notice from the truth conditions, or from the reduction schemes above and properties of $\Box$ and K_i, that the following axiom schemas are valid (among others):

$$\begin{array}{rl} \mathrm{K}_i \, p \rightarrow \widehat{\mathrm{K}}_i^j \, p & \quad \text{(but not the converse)} \\ \mathrm{K}_i \, \mathcal{O}_j \, p \rightarrow \widehat{\mathrm{K}}_i^j \, p & \quad \text{(but not the converse)} \\ \mathcal{O}_j \, p \rightarrow \widehat{\mathrm{K}}_i^j \, p & \quad \text{(but not } \mathcal{O}_j \, p \rightarrow \mathrm{K}_i \, \mathcal{O}_j \, p) \end{array}$$

This seems intuitively correct. If one restricts attention to states in which j is functioning correctly, i 'knows' at least as much as when all states, j-green and j-red, have to be considered (first of the axioms). And if i knows that p holds in all states where j is functioning correctly, i.e. $\mathrm{K}_i \, \mathcal{O}_j \, p$ holds, then surely also $\widehat{\mathrm{K}}_i^j \, p$; on the other hand, there could be things p that i 'knows' on the assumption that j is functioning correctly that do not hold in all j-correct states: $\widehat{\mathrm{K}}_i^j \, p$ should not imply $\mathrm{K}_i \, \mathcal{O}_j \, p$. Of course, to be really useful, the question is not just whether $\widehat{\mathrm{K}}_i^j \, p$ holds but whether i can determine this, i.e. whether $\mathrm{K}_i \, \widehat{\mathrm{K}}_i^j \, p$ holds. But notice: $\widehat{\mathrm{K}}_i^j \, p \rightarrow \mathrm{K}_i(\mathbf{g}_j \rightarrow p)$ (by definition), $\mathrm{K}_i(\mathbf{g}_j \rightarrow p) \rightarrow \mathrm{K}_i \, \mathrm{K}_i(\mathbf{g}_j \rightarrow p)$ (by property 4 of K_i), and $\mathrm{K}_i \, \mathrm{K}_i(\mathbf{g}_j \rightarrow p) \rightarrow \mathrm{K}_i \, \widehat{\mathrm{K}}_i^j \, p$ (by definition). Since we also have $\mathrm{K}_i \, p \rightarrow p$, we have the following valid axiom:

$$\widehat{\mathrm{K}}_i^j \, p \leftrightarrow \mathrm{K}_i \, \widehat{\mathrm{K}}_i^j \, p \qquad (\text{all } i \in A),$$

which seems very satisfactory.

As for the relationship between $\mathcal{O}_j \, \mathrm{K}_i$ and $\widehat{\mathrm{K}}_i^j$, various interactions can readily be determined, such as the following:

$$\mathcal{O}_j \, \mathrm{K}_i \, p \rightarrow \widehat{\mathrm{K}}_i^j \, \mathrm{K}_i \, p$$
$$\mathcal{O}_j \, \mathrm{K}_i \, p \rightarrow \widehat{\mathrm{K}}_k^j \, \mathrm{K}_i \, p \qquad \text{(any } k \in A\text{)}$$

We would also like to be able to give a complete characterisation of the language of $\mathcal{O}_i$, K_i, and $\widehat{\mathrm{K}}_i^j$ in terms of Kripke frames. It is worth noting that the doubly-indexed operator $\widehat{\mathrm{K}}_i^j$ would be interpreted on the intersection of the relations corresponding to $\mathcal{O}_i$ and K_i. Providing axiomatisations for operators defined on intersections of relations is non trivial. One of the cases that are better known from the literature is the case of distributed knowledge [FHV92, HM92]. Here it is known that one can obtain a complete axiomatisation for a multi-agent epistemic language with distributed knowledge D by adding S5 axioms to the operator D and taking the axiom $\vee_{i=1,\ldots,n} K_i p \rightarrow Dp$. The complication of the current setting over distributed knowledge is twofold. For the case of distributed knowledge, first all the relations have the same properties; second they are equivalence relations. For the case under consideration here, while it is easy to see that the intuitively corresponding axiom:

$$\mathcal{O}_j \, p \vee \mathrm{K}_i \, p \rightarrow \widehat{\mathrm{K}}_i^j \, p \tag{1}$$

is valid on the relevant semantic structures, one cannot apply the results presented in the literature. [FHV92] uses a reduction to equivalence Kripke trees which cannot be applied here because R_i is not an equivalence relation. The proof used in [HM92] can be used for relations that are not necessarily equivalence relations, but the authors do assume that the relations from which the intersection is taken have the same properties. Still, we are hopeful that completeness can be proven by extending the rewriting technique used in [HM92], and it is reasonable to expect to have a logic whose fragments are KD45 for each $\mathcal{O}_i$ component, S5 for each K_i component, K45 for each $\widehat{\mathrm{K}}_i^j$ component and the interaction axiom (1).

6. Applications

One of our motivations for developing the theoretical constructs presented so far is the following. We want to be able to examine what (epistemic) properties of a system hold when agents conform to their specifications/protocols, and to determine which of these (epistemic) properties are retained and which are compromised when agents fail to conform to their specification. So: we want to be able to set up a model of a system which allows for and covers the case where agents' behaviours may deviate from what is specified. Generally only a subset of an agent's possible behaviours will be regarded as acceptable/permitted ('green'); the model will also specify the behaviours that agents should exhibit when in non-acceptable/non-permitted ('red') states. We are interested then in determining what (epistemic) properties of the system hold if we restrict attention to the case where all agents conform to specification, i.e., to the case where all states are green. We are interested in determining what (epistemic) properties of the system hold on the assumption that agents i, j, ... , conform to their specifications. And further, once we allow for the possibility that an agent's behaviour may deviate from its specification, it is natural to consider adding extra *control* components to the system, either as part of the environment or by introducing additional agents whose function is to constrain the behaviours of one or more agents to force, to some degree or other, compliance with the specifications. In that case we want to be able to investigate what (epistemic) properties of this extended system are recovered—and to what extent they in turn depend on the assumption that the control agents conform to their specifications.

One example for which we have been able to carry out such an analysis is the bit transmission protocol (BTP). The BTP involves two agents, one sender S, and one receiver R, trying to communicate over a faulty channel. The channel is faulty in that it may non-deterministically deliver a message sent on it or lose it altogether. It may deliver a message sent in one direction whilst simultaneously losing one sent in the other direction. It may lose both messages, or neither. S would like to send a bit of information to R, and would like to know for sure when R has received the bit. According to the most common protocol, sender S sends the bit until it receives an acknowledgement from R; R remains silent until it receives the bit and it then sends acknowledgements to S forever. This setting has been modelled in the interpreted systems framework [FHMV95]. The set of local states for sender S, $L_S = \{0, 1, (0, ack), (1, ack)\}$ represents configurations in which the value of the bit is either 0 or 1, paired with whether or not the bit has been received. Similarly R can be modelled by $L_R = \{\epsilon, 0, 1\}$, in which ϵ represents the configuration in which no bit has been received by R and 0, 1, local states in which the bit has been copied successfully by R. Given this, it is possible fully to specify functioning protocols for sender and receiver by defining functions for each local state what actions should be performed by that agent. For example, according to the protocol if S is in state 0 it should send the value 0 across the channel to R. We refer to [FHMV95] for more details.

What is important in this semantical construction is that epistemic properties of the system can be analysed by using the formal tools of interpreted systems *generated by such protocols*. In particular, under the assumption of fairness (i.e., assuming that the environment will not be faulty forever), one can show that the following properties are valid (true at all global states of the system):

$$\models \quad \mathbf{recbit} \rightarrow \big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \tag{2}$$

$$\models \quad \mathbf{recack} \rightarrow \mathbf{recbit} \tag{3}$$

which capture our intuition about the model. From these it follows (in $S5_n$) that:

$$\models \quad \mathbf{recack} \rightarrow \big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \tag{4}$$

$$\models \quad \mathbf{recack} \rightarrow \mathrm{K}_S\big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \tag{5}$$

$$\models \quad \mathbf{recack} \wedge (\mathbf{bit} = \mathbf{0}) \rightarrow \mathrm{K}_S\,\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \tag{6}$$

(and similarly for the case $(\mathbf{bit} = \mathbf{1})$). So, if an ack is received by S, then S is sure that R knows the value of the bit. Intuitively this represents the fact that although the channel is potentially faulty, if messages do manage to travel back and forth the protocol is strong enough to eliminate any uncertainty in the communication. Whereas, for example,

$$\not\models \; \mathbf{recack} \wedge (\mathbf{bit} = \mathbf{0}) \rightarrow \mathrm{K}_R\,\mathrm{K}_S\,\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0})$$

In the basic version of the BTP discussed above, the environment's faults are treated as a kind of uncertainty in the system and modelled by non-deterministic evolution. But there are more fundamental flaws that we may be interested in examining, in particular, flaws that pertain to the agents' behaviour rather than the environment's. Suppose, to take a concrete example, that R may fail to comply to the functioning behaviour specified in the protocol in that it may incorrectly send an acknowledgement when it has not yet received the bit (i.e., while in state ϵ). We can model this situation semantically in deontic interpreted systems by extending the set of possible local states for R and labelling as faulty ('red') those in which the acknowledgement is incorrectly sent by R. If we carry out an analysis of the epistemic properties that result from this modification (the details may be found in [LS02]) we discover that the key property is no longer valid

$$\not\models \mathbf{recack} \rightarrow \mathbf{recbit}$$

and that (among other things):

$$\not\models \mathbf{recack} \rightarrow K_S(K_R(\mathbf{bit} = \mathbf{0}) \vee (K_R(\mathbf{bit} = \mathbf{1}))).$$

However, using the operator $\mathcal{O}_R$ which represents what holds in states where R is operating correctly, we do have the following:

$$\models \quad \mathcal{O}_R(\,\mathbf{recack} \rightarrow \mathbf{recbit}\,) \tag{7}$$

$$\models \quad \mathcal{O}_R(\,\mathbf{recack} \rightarrow \big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big)\,) \tag{8}$$

Furthermore, and more interesting, if S makes the assumption of R's correct functioning behaviour, then, upon receipt of an acknowledgement, it would make sense for it to assume that R does know the value of the bit. The operator $\widehat{\mathrm{K}}_i^j$ representing "knowledge under the assumption of correct behaviour' models this notion. For example, the following are both valid on the models:

$$\models \quad \mathbf{recack} \rightarrow \widehat{\mathrm{K}}_S^R\big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \tag{9}$$

$$\models \quad \mathbf{recack} \wedge (\mathbf{bit} = \mathbf{0}) \rightarrow \widehat{\mathrm{K}}_S^R\,\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \tag{10}$$

It is possible and instructive to analyse other variants of this scenario. Suppose for example that R were to fail in the following (different) way: if the bit is received it may or may not send an acknowledgement. Intuitively this sort of failure is less problematic than the case where R sends incorrect acknowledgements before receiving the bit, since it does not compromise communication. Again, this analysis can be carried out in deontic interpreted systems: indeed, in this context one can explore *error correcting protocols* for R — essentially it makes sense to impose that from a red state R can recover to a green state just by sending the acknowledgement. It can be checked that an error correcting protocol of this sort does guarantee that formula (1) above is valid.

More complex variations can be devised. For example one can introduce a third agent, a controller C, whose task is to take preemptive action before critical failures occur. Suppose for example that controller agent C monitors the environment (according to the spirit of interpreted systems, the internal local state of agent R is invisible to C) and blocks any attempts by R to send incorrect acknowledgements over the channel. Constructing the model for this scenario, we find that the following hold:

$$\begin{aligned} &\models \quad \mathbf{recbit} \rightarrow \big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \\ &\models \quad \mathcal{O}_C(\mathbf{recack} \rightarrow \mathbf{recbit}) \end{aligned}$$

from which follows, by the same derivations as earlier:

$$\begin{aligned} &\models \quad \mathcal{O}_C(\mathbf{recack} \rightarrow \big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big)) \\ &\models \quad \mathbf{recack} \rightarrow \widehat{\mathrm{K}}_S^C\big(\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \vee \mathrm{K}_R\,(\mathbf{bit} = \mathbf{1})\big) \\ &\models \quad \mathbf{recack} \wedge (\mathbf{bit} = \mathbf{0}) \rightarrow \widehat{\mathrm{K}}_S^C\,\mathrm{K}_R\,(\mathbf{bit} = \mathbf{0}) \end{aligned}$$

Here, as one might expect, communication between the sender and receiver is re-established, under the assumption that controller C behaves correctly according to its specification.

For details of how these models are constructed and analysed, the interested reader is referred to [LS02].

7. Conclusions

In this paper we have explored the notion of deontic systems of global states and axiomatised the resulting semantics. Apart from the technical issues dealt with in the paper, we have tried to motivate the suggestion of exploring grounded semantics (interpreted systems or otherwise) to give an interpretation to deontic concepts when these are aimed at computer science applications.

Various technical issues seem to be worth exploring further. In particular the question of finding a complete axiomatisation for the operator $\widehat{\mathrm{K}}_i^j$ discussed above seems promising. Once this is achieved it will be interesting to explore the notion of common knowledge with respect to deontic states and groups of agents.

The framework as presented here is very general. Although we have tended in the discussion to associate the green label with 'correct' or 'allowed' functioning and red with 'incorrect' or 'disallowed', nothing in the formal development relies on this particular reading. The green/red labelling of states could be used just as well to model other distinctions: green could be used to pick out, for instance, the normal, non-exceptional states from the abnormal, exceptional ones. The $\widehat{\mathrm{K}}_i^j$ operator would then express what agent i knows on the assumption that agent j is not in an unusual, exceptional state. Perhaps we have a means of modelling defeasible knowledge. This remains to be investigated.

We have tried to apply the formal machinery to a standard example in the literature of distributed computing - the bit transmission problem. We have analysed how violations and correct functioning behaviour of parts of the system can be represented in deontic interpreted systems, and how the effects of introducing additional control components into the system can be determined. Of course the example is trivial compared to the kind of complex MAS applications now being deployed. Nevertheless, the fact that we have managed to analyse an example in detail using the formal machinery (further details can be found in [LS02]) encourages us to believe that the methods can be applied to more complex examples. Manual calculations of even small examples, however, are tedious and prone to error. In order to apply the formal machinery to more complex and realistic examples, we have been experimenting with the use of model checking tools. Specifically, we have used a model checker for temporal logic to compute the set of all possible runs of a system where the protocols for the agents are given (as in the BTP), and we then use a model verifier to check the validity of particular epistemic formulas on the resulting models. We refer the interested reader to [LRS02].

We are aware that the analysis of protocol violations and message transmission failures is a major research issue in the area of distributed algorithms. We are interested in investigating how the results of this paper compare with the methods that have been developed there.

The formal language we have used in this paper has no temporal constructs. This is an aspect of this work that we are planning to address.

Acknowledgements The authors are grateful to Ron van der Meyden for his comments on an earlier draft of this paper. This research was supported by EU Framework ALFEBIITE IST-1999-10298.

Appendix

We report two Lemmas used in the proofs of Section 3.

Notice that seriality is required for the first but not for the second of these two lemmas.

LEMMA 6. *Let $F = (W, R_1, \ldots, R_n)$ be a transitive, i-j Euclidean and serial frame. Then for every $i \in A$, and for all $w, w', w'' \in W$, if wR^*w' and $w'R_iw''$ then wR_iw''.*

PROOF. wR^*w' means there is a path $wR_{i_1}v_1, v_1R_{i_2}v_2, \ldots, v_{k-1}R_{i_k}w'$. We have to show $wR_{i_1}v_1, v_1R_{i_2}v_2, \ldots, v_{k-1}R_{i_k}w', w'R_iw''$ implies wR_iw''. The proof is by induction on the length k of the path. For the base case ($k = 0$), $w = w'$ and the result holds trivially. For the inductive step, suppose the result holds for all paths of length less than k. It suffices to show $v_{k-1}R_iw''$, for then we have a path of length $k-1$ from w to v_{k-1} and by the inductive hypothesis it follows that wR_iw''. To show $v_{k-1}R_iw''$, consider $v_{k-1}R_{i_k}w'$. Suppose first that $i_k = i$. Then we have $v_{k-1}R_iw'$ and $w'R_iw''$, and so $v_{k-1}R_iw''$ by transitivity of R_i. Suppose $i_k \neq i$. Then because R_i is serial, there exists a point w''' such that $v_{k-1}R_iw'''$. We must have also $w'R_iw'''$ since the frame is i-j Euclidean, and further, $w'''R_iw''$ because $w'R_iw''$ and R_i is Euclidean. So now we have $v_{k-1}R_iw'''$ and $w'''R_iw''$, and by transitivity of R_i, $v_{k-1}R_iw''$ as required to complete the proof. ∎

LEMMA 7. *Let $F = (W, R_1, \ldots, R_n)$ be an i-j Euclidean frame. Then for every $i \in A$, and for all $w, w', w'' \in W$, if wR^*w' and wR_iw'' then $w'R_iw''$.*

PROOF. The proof is again by induction on the length of the path in wR^*w'. Suppose $w = w'$ (base case). Then the result holds trivially. Suppose $w \neq w'$. Then there is a path $wR_{i_1}v_1, v_1R_{i_2}v_2, \ldots, v_{k-1}R_{i_k}w'$, and for the inductive step we have to show $wR_{i_1}v_1, v_1R_{i_2}v_2, \ldots, v_{k-1}R_{i_k}w', w'R_iw''$ implies wR_iw'', assuming the result holds for all paths of length less than k. wR_iw'' implies v_1R_iw'' since the frame is i-j Euclidean. Now we have a path v_1R^*w' of length $k-1$ and v_1R_iw'', so $w'R_iw''$ follows by the inductive hypothesis. ∎

References

[And58] A. R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 58:100–103, 1958.

[Åqv67] L. Åqvist. Good samaritans, contrary-to-duty imperatives, and epistemic obbligations. *NOUS*, 1:361–379, 1967.

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2001.

[Che80] B. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.

[FHMV95] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, 1995.

[FHV92] R. Fagin, J. Y. Halpern, and M. Y. Vardi. What can machines know? On the properties of knowledge in distributed systems. *Journal of the ACM*, 39(2):328–376, April 1992.

[Gab98] D. Gabbay. *Fibring Logics*. Oxford University Press, 1998.

[GMP92] J. Glasgow, G. MacEwen, and P. Panangaden. A logic for reasoning about security. *ACM Transactions on Computer Systems*, 10(3):226–264, August 1992.

[Gol92] 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.

[GS98] D. Gabbay and V. Shehtman. Products of modal logics, part 1. *Logic Journal of the IGPL*, 6(1):73–146, 1998.

[HC96] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, New York, 1996.

[HM90] J. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990. A preliminary version appeared in *Proc. 3rd ACM Symposium on Principles of Distributed Computing*, 1984.

[HM92] W. van der Hoek and J.-J. Ch. Meyer. Making some issues of implicit knowledge explicit. *International Journal of Foundations of Computer Science*, 3(2):193–223, 1992.

[JS93] A. J. I. Jones and M. J. Sergot. *Deontic Logic in Computer Science: Normative System Specification*, chapter 12: On the Characterisation of Law and Computer Systems: The Normative Systems Perspective. Wiley, 1993.

[Kri59] S. A. Kripke. Semantic analysis of modal logic (abstract). *Journal of Symbolic Logic*, 24:323–324, 1959.

[KW91] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56(4):1469–1485, 1991.

[LMR00] 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.

[LR98] A. Lomuscio and M. Ryan. On the relation between interpreted systems and Kripke models. In M. Pagnucco, W. R. Wobcke, and C. Zhang, editors, *Agent and Multi-Agent Systems - Proceedings of the AI97 Workshop on the theoretical and practical foundations of intelligent agents and agent-oriented systems*, volume 1441 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, May 1998.

[LRS02] A. Lomuscio, F. Raimondi, and M. Sergot. Towards model checking interpreted systems. In *Proceedings of Mochart — First International Workshop on Model Checking and Artificial Intelligence*, 2002.

[LS02] A. Lomuscio and M. Sergot. Violation, error recovery, and enforcement in the bit transmission problem. In *Proceedings of DEON'02*, London, May 2002.

[Woo00] M. Wooldridge. Computationally grounded theories of agency. In E. Durfee, editor, *Proceedings of ICMAS, International Conference of Multi-Agent Systems*. IEEE Press, 2000.

ALESSIO LOMUSCIO
Department of Computer Science
King's College London
Strand
London WC2Z 2LS, UK
alessio@dcs.kcl.ac.uk

MAREK SERGOT
Department of Computing
Imperial College
180 Queen's Gate, London SW7 2BZ, UK
mjs@doc.ic.ac.uk