On multi-agent systems specification via deontic logic

Alessio Lomuscio Marek Sergot

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

Abstract. We investigate an extension of the formalism of interpreted systems by Halpern and colleagues to model 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, $\mathrm{KD45}_n^{i\text{-}j}$, is a stronger version of KD, the system often referred to as Standard Deontic Logic. We discuss these issues and present some simple examples.

1 Introduction

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.

Somewhat unsurprisingly, the advent of the intelligent agents paradigm in distributed computing, together with the tremendous growth of the Internet medium, has widened this gap even further. On the one hand, the intrinsic complexity of the capabilities required by agents has encouraged theoreticians to consider logics for intensional concepts developed by philosophers; on the other hand, the commercialisation of the Internet together with the emphasis on timed product delivery from software houses has progressively led practitioners to ignore the results being produced by theorists of multi-agent systems (MAS).

Still, if MAS (see [18] for a recent introduction) are to fulfil the promises of providing the architectural foundations of increasingly critical and complex Internet applications, such as automatic negotiation in e-commerce, remote control of safety critical systems, etc, then there is clearly a need for bridging the gap between theory and practice. In our view, this is particularly relevant with respect to the issues relating to specification and verification of MAS.

One of the suggestions that have been put forward [17] (and that we fully endorse) 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 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, 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 [8]. 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 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), 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. Notwithstanding this, no clear correspondence can be drawn between a Kripke model and a distributed computing system; in particular, it is a matter of debate whether or not a correspondence can be drawn between the notion of world in a Kripke model and that of state of a computational 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 [2] 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 quite 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 logics 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 [2] 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. Once this task is accomplished, we axiomatise this resulting semantical class and draw correspondences between the obtained results and some well-known in the literature of deontic logic.

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. In Section 4 we comment on the results of the paper and draw attention to issues for further work. We conclude in Section 5.

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.*

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

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 $\mathcal{O}_i$: $\mathcal{P}_i\, \varphi =_{def} \neg\, \mathcal{O}_i\, \neg\varphi$. Intuitively, $\mathcal{P}_i\, \varphi$ stands for "in some of the states in which agent i operates correctly φ holds", or "φ holds in some of the correct functioning alternatives of agent i". 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".

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 KT45$_n$ we denote the normal logic obtained by considering axioms K, T, 4, and 5 for n agents (see [2] for more details). In the following we assume familiarity with basic modal logic techniques and results. We refer the reader to [7, 13] for more details.

2.2 Deontic interpreted systems

Interpreted systems were originally defined by Halpern and Moses [5], and their potentiality later presented in greater detail in [2]. 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 $L_1, \ldots, L_n$ 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 $l_1, l_1', l_2, l_2', \ldots$. Elements of L_e will be denoted by $l_e, l_e', \ldots$.

Definition 2 (System of global states). *A* system of global states for n agents S *is a non-empty subset*[1] *of a 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 [2] 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.

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 3 (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.

[1] The case of the full Cartesian product was analysed in [9].

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 4 (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 [10]). In this case, the relations on the generated Kripke frame are equivalence relations; hence (see [15, 2]) 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 axiomatize deontic systems of global states using the languages defined in Definition 1, 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 5 (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 deontic 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 $F(DS, \pi)$.

Definition 6 (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:

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

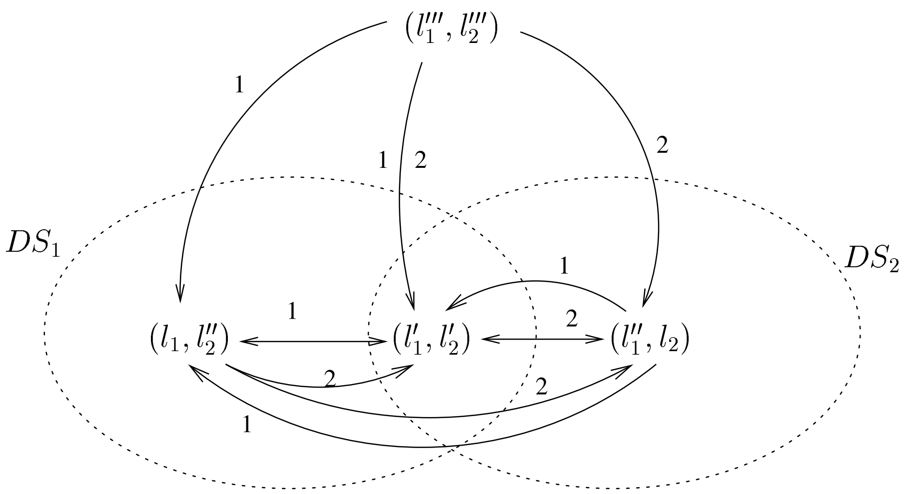

Fig. 1. An example of deontic system and its generated frame. In the example above the environment is not considered and the local states for the agents are composed as follows. Agent 1: $L_1 = \{l_1, l_1', l_1'', l_1'''\}, G_1 = \{l_1, l_1'\}$. Agent 2: $L_2 = \{l_2, l_2', l_2'', l_2'''\}, G_2 = \{l_2, l_2'\}$. $DS = \{(l_1', l_2'), (l_1, l_2''), (l_1'', l_2), (l_1''', l_2''')\}$. 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.

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 7 (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 7 it follows that this analysis can be carried out on the class of the generated frames.

3 Axiomatisation of deontic interpreted systems

In this section[2] we study deontic systems of global states from the axiomatic point of view. An immediate consideration comes from the following.

[2] Proofs of many of the results in this section are not included in this version. We refer the interested reader to [11] for details.

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

This observation leads immediately to the conclusion that the logic of deontic systems of global states must be at least as strong as $\mathrm{KD45}_n$, 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 $\mathrm{KD45}_n$. In order to see this, we need to introduce a few semantic structures.

3.1 Some secondary properties of Kripke frames

In order to prove 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 [1, p92].)

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

Definition 8 (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.

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 9 (i-j Euclidean frame). *A frame $F = (W, R_1, \ldots, R_n)$ is* i-j Euclidean *if 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:

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

Lemma 3. *A frame F is i-j Euclidean if and only if $F \models 5^{i\text{-}j}$.*

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

Lemma 4. *If a frame F is secondarily universal then it is also i-j Euclidean.*

Lemma 5. *Let $F = (W, R_1, \dots, R_n)$ 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 $w'R_iw''$, where R^* is the reflexive transitive closure of $\cup_{i=1}^n R_i$.*

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 1. *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.*

Theorem 2. *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 1. 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 3. *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 $DS = \{(w, w, \dots, w) \mid w \in W\}$. For simplicity we use the shortcut $(w, w, \dots, w) = [w]$. The frame generated by DS is defined (see Definition 5) by $F(DS) = F' = (W', R'_1, \dots, R'_n)$, 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 $p([w]) = w$.

- 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 sufficient, 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 4. *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 2, there exists a serial, secondarily universal model $M = (F, \pi)$ such that $M \not\models_w \varphi$, for some $w \in W$. By Theorem 3 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 consideration on p-morphisms, 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:

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

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:

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

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, then axiom 4

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

and axiom 5

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

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 $g_i : L_i \to 2^{L_i}$ 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 5); $\mathcal{O}_i\, \mathcal{O}_i\, p \rightarrow \mathcal{P}_i\, \mathcal{O}_i\, p$ (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 $2n$ distinct modalities in the logic $\mathrm{KD45}_n^{i\text{-}j}$.

4.2 A notion 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 several 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:

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

There are several possible definitions of G, depending on the notion of global correctness we wish to model:

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:

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

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 Further work and conclusions

In this paper we have explored the notion of deontic systems of global states and axiomatised the resulting semantics. This line of work lies within a growing body of research that deals with the issue of axiomatising Multi-Agent systems defined on computationally grounded semantics, in the sense of [17]. This approach was suggested in the influential work by Halpern and colleagues [3] in which the authors studied the axiomatic properties of different classes of interpreted systems. Recently, some works have tried to relate these issues to semantics with a finer level of detail [9]. Of particular importance in this line of research is the issue of proving completeness not just with respect to an abstract class of Kripke frames, but with respect to an automata like model of a distributed system, as done in this present paper. This line of analysis opens up interesting avenues with respect to performing model checking directly on these semantical classes in the spirit of [6].

Various technical issues seem to be worth exploring further. In particular the question of incorporating mentalistic attitudes on the semantics we used here seems interesting. We have already undertaken some analysis for the case of knowledge [12]. A similar analysis for the case of belief could perhaps lead to interesting remarks regarding an axiomatisation of the *s*incerity condition of the *inform* communicative act of the FIPA specification [4]. As the current specification stands, it states that a pre-condition for an *inform(s,r, p)* act is that sender s believes that p is the case. This strong condition has led to some criticism. Indeed, one can imagine various different scenarios in which deception is a technique that may be used by some agent. For example in a negotiation scenario it does not seem realistic to insist that the agents exchange the true limit price for the goods or service they are negotiating upon. In this context, it seems of some interest to explore the possibility of modifying the sincerity condition into $\mathcal{O}_s\, B_s p$. In this way, agent s would violate its own (perhaps certified by a third party) functioning protocol by sending an untruthful information, but it would still be practically possible for it to do so (and possibly face the consequences of this later on). The implications of this seem worth exploring; we do not carry out this exercise here.

Another line of work which we have recently started pursuing concerns the evaluation of the applicability of the present framework to modelling real scenarios, where one is typically required to reason about correct functioning behaviour of the agents and of the system as a whole. We briefly report only two cases here.

Agents communicating over a channel. Consider an operating systems scenario in which two agents communicate over a channel. Agent 1 is a monitor, agent 2 is a cron job. Let us assume that agent 2 occasionally sends information to agent 1 via the channel. Each agent may or may not be in a faulty state. We represent the agents' local states as $L_i = \{+, -\}, i = \{1, 2\}$, where the symbol $+$ represents a state of correct functioning behaviour, whereas $-$ stands for a faulty state; we capture the functioning behaviour of the communication channel in the same way. Clearly the set of states of the system consists of all the tuples $DS = \{(l_1, l_2, l_e)\}$ in which all the components vary on $\{+, -\}$. Consider now the atoms $P = \{p, q\}$ where p stands for "agent 1 is alive, i.e. listening", and q "agent 2 is sending messages". We would like to consider formulas like:

$$\mathcal{O}_1\, p \wedge \neg\, \mathcal{O}_2\, q \wedge \neg\, \mathcal{O}_2 \neg q,$$

representing models in which agent 1 is supposed to be monitoring the channel (indeed we would model p as being true only in the global states where agent 1 is functioning correctly), whereas agent 2 may or may not be sending messages. Note that, when taking the bird's eye view, there is nothing problematic about the reading of formulas such as $\mathcal{O}_2\, p$ and $\mathcal{O}_1\, q$. Consider now a notion of global correctness corresponding to the environment, i.e. the communication channel in our scenario, working correctly. One may want to check whether or not particular models satisfy the formula:

$$\mathcal{O}(q \rightarrow p),$$

i.e., models in which whenever the communication channel is operational, it follows that all messages are successfully received. Note that this does not hold in the case of the full Cartesian product described here. An interesting question is what sort of additional enforcement mechanism it is possible to impose so that some wanted property, like the one above, is exhibited.

Financial liquidity. Consider the case of a financial institution in which the current accounts of customers are modelled as agents. Agents can perform different actions in the financial environment (arranging overdraft etc.) and perform actions affecting other agents (bank transfer etc.), so it is meaningful to take this perspective. Let us suppose that the agents' local states simply represent the balance of the account. We can introduce a notion of individual violation for the agents representing that the account must not be overdrawn, i.e. the set of red states for every agent is the one containing, say, all negative natural numbers. An agent being overdrawn clearly poses no risk to the general financial health of the institution, but the overall balance of the set of agents surely does. It is easy to see how to model these issues in terms of a global/local violations as described above by considering an appropriate set of propositional variables. But further, one would want to devise an enforcement mechanism that forces no global violation (i.e. all the agents collectively being in the black) while allowing for individuals to violate their individual constraints.

The examples above show that while the formalism seems adequate for representing local and global violations, it needs to be made stronger to represent more interesting concepts, such as regulation, and various kinds of enforcement/compliance mechanisms. Indeed from the syntactical/proof-theoretical point of view we have no more than a stronger version of Standard Deontic Logic, and therefore we are only likely to inherit its well known limitations, such as it inability to model contrary to duty structures [16]. We see this paper as a starting point. Our current plan is to investigate the extent to which generalisations of the present formalism to incorporate temporal evolution and action can be used to represent enforcement/compliance mechanisms.

Acknowledgements

This work was supported by the EU-funded FET project *ALFEBIITE* (IST-1999-10298).

References

1. B. Chellas. *Modal Logic*. Cambridge University Press, Cambridge, 1980.
2. R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, 1995.
3. 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.

4. FIPA: Foundation for intelligent physical agents. http://www.fipa.org.
5. 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.
6. J. Halpern and M. Vardi. *Model checking vs. theorem proving: a manifesto*, pages 151–176. Artificial Intelligence and Mathematical Theory of Computation. Academic Press, Inc, 1991.
7. G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, New York, 1996.
8. S. A. Kripke. Semantic analysis of modal logic (abstract). *Journal of Symbolic Logic*, 24:323–324, 1959.
9. 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.
10. 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.
11. A. Lomuscio and M. Sergot. Investigations in grounded semantics for multi-agent systems specifications via deontic logic. Technical report, Imperial College, London, UK, 2000.
12. A. Lomuscio and M. Sergot. Extending interpreting systems with some deontic concepts. In J. van Benthem, editor, *Proceedings of TARK 2001*, pages 207—218, San Francisco, CA, July 2001. Morgan Kauffman.
13. J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambdridge University Press, 1995.
14. J.-J. Ch. Meyer and R. J. Wieringa. *Deontic Logic: A Concise Overview*, chapter 1, pages 3–16. Wiley Professional Computing Series. John Wiley and Sons, Chichester, UK, 1993.
15. S. Popkorn. *First Steps in Modal Logic*. Cambridge University Press: Cambridge, England, 1994.
16. H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57(1):91–115, 1996.
17. M. Wooldridge. Computationally grounded theories of agency. In E. Durfee, editor, *Proceedings of ICMAS, International Conference of Multi-Agent Systems*. IEEE Press, 2000.
18. M. Wooldridge. *Reasoning about rational agents*. MIT Press, July 2000.