Extending interpreted systems with some deontic concepts*

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 interpreted systems to model correct functioning behaviour of agents and of the system as a whole. We combine this notion with the standard epistemic notions defined on interpreted systems to provide a formalism to reason about knowledge that agents are permitted to hold under ideal functioning circumstances. We then extend this by introducing a doubly-indexed operator representing knowledge that an agent would have if it were operating under the assumption that a group of agents is functioning as intended. We investigate the completeness problem for the first formalism and discuss the issue for the more general one.

1 Introduction

The area of modal epistemic logic, developed into its modern form by the work of Hintikka [Hin62, HM92a] in philosophical logic, has found promising applications in computer science [FHMV95, MH95], and economics [Aum76]. Technically the work revolves around a family of modal logics that can be used to give a bird's eye view of the knowledge properties of a multi-agent system.

While most of the knowledge representation literature concerns *explicit knowledge*, i.e., knowledge that the agents themselves are aware of and which informs their actions and decisions, reasoning about knowledge of agents or processes from an observer's perspective (the bird's eye view) is essential in other areas of computer science. Only in this way can we reason about the information that agents in the system have *in principle* at their disposal; for example, in cryptography it is crucial to verify whether or not an agent has enough information to decipher a message regardless of whether or not that agent has in fact been able to decipher the message. Evidence of the interest in modal logic as a specification language for epistemic notions can be found in the multitude of logics that have been discussed, and in the different semantics (interpreted systems [HM90], contexts [FHMV95], environments [Mey96]) that have been proposed to model different grains of detail of communication processes, protocols, etc.

One of the most developed parts of these studies involves the combination between knowledge and other modal operators, notably knowledge and belief [Hoe93], and knowledge and time [HMV97]. This has brought about a greater understanding of (a-)synchronicity, recall capabilities, and interaction between knowledge states and other aspects of agency. Much of this literature is not only concerned with axiomatisations of the knowledge properties of particular multi-agent systems but most interestingly with the notion of *protocol*. In particular, it is of interest to reason about the knowledge properties of a multi-agent system that are enforced by a particular protocol. Consider for example agents that start in a hypercube configuration (i.e., one in which all the local states for the agents are equally possible); in this case it can be shown that if the protocol enforces synchronous broadcast the knowledge properties of the system are captured by the logic S5WD$_n$ (see [LMR00] for details).

While it is worthwhile to study protocols enforcing particular epistemic states from an axiomatic point of view, it is also of interest to analyse systems in which there is no guarantee that the intended protocol

* The authors acknowledge support from EU project ALFEBIITE, IST-1999-20298.

will be followed by all the agents. For example, this is an issue in safety critical systems where one wants to reason about the properties of the individual agents and of the system as a whole not only when functioning as intended but also, and equally importantly, when *not* doing so. The distinction between actual and intended or ideal functioning behaviour has been discussed in possible applications of deontic logic (e.g., see [MW93]). Still, to the best of our knowledge, these works have stopped short of investigating the knowledge properties of the agents that arise when a violation in the protocol has or has not occurred.

In this article we would like to take a first step towards developing logical formalisms for reasoning about knowledge properties relating to intended and non-intended functioning behaviour of the agents. As ever there exist two approaches in which to carry out this exercise: syntactical and semantical. From a syntactical point of view one may consider combinations [BdR97] of deontic and epistemic operators, and study logics that include KD_n for the deontic component together with $\mathrm{S5}_n$ for the epistemic one. This is likely to give rise to interesting axiomatisation problems, and it is a worthwhile exercise, but the axiomatisations that one would obtain may not be easy to relate to intuitive classes of computational processes, especially if carried out with respect to classes of Kripke frames.

Alternatively, one may start from the semantics, and in particular from the intuitive framework of interpreted systems as defined by Halpern et al. in [FHMV95], encode the concepts of ideal/correct behaviour there, and study the interaction of these with the usual epistemic notions of epistemic constructions. In this paper we run an exercise along these lines. We start with the basic notion of interpreted system, and show how it can be trivially adapted to represent some issues normally addressed in deontic logic. 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 will present a complete axiomatisation of the concept of *ideal (or normal) functioning behaviour* of an agent, and of a system of agents. Having done so, we will introduce the concept of the *knowledge* that an agent is *permitted to have* (again with respect to an ideal functioning protocol), and of the knowledge that an agent has *on the assumption* that components of the system are functioning correctly according to their protocols, and we will study a few different ways in which this can be encoded in the formalism.

This paper is organised as follows. 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. Next in Section 4 we use these results to incorporate knowledge, and reason about knowledge under correct behaviour. We conclude in Section 5.

2 Deontic interpreted systems

2.1 Syntax

We will start by analysing a simple indexed deontic language, and later extend it with an indexed epistemic operator. For the moment 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 $\mathcal{O}_i$: $\mathcal{P}_i\,\varphi =_{def} \neg\,\mathcal{O}_i\,\neg\varphi$. 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. We use, and assume knowledge of, standard notions and constructions of Kripke semantics and modal logic systems. See [HC96, FHMV95] for details.

2.2 Deontic interpreted systems

Interpreted systems were originally defined by Halpern and Moses [HM90], and their potentiality later presented 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.

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 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 [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, and work only on states.

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. A different but interesting approach is to label runs instead of states. In this way we would be able to reason about ideal, or normal runs as opposed to non-ideal/non-normal runs. We do not explore these ideas here.

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, e.g., rationality if the agents are players in a game theoretical setting.

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 [LR98]). In this case, the relations on the generated Kripke frame are equivalence relations; hence (see [Pop94, FHMV95]) the logic resulting by

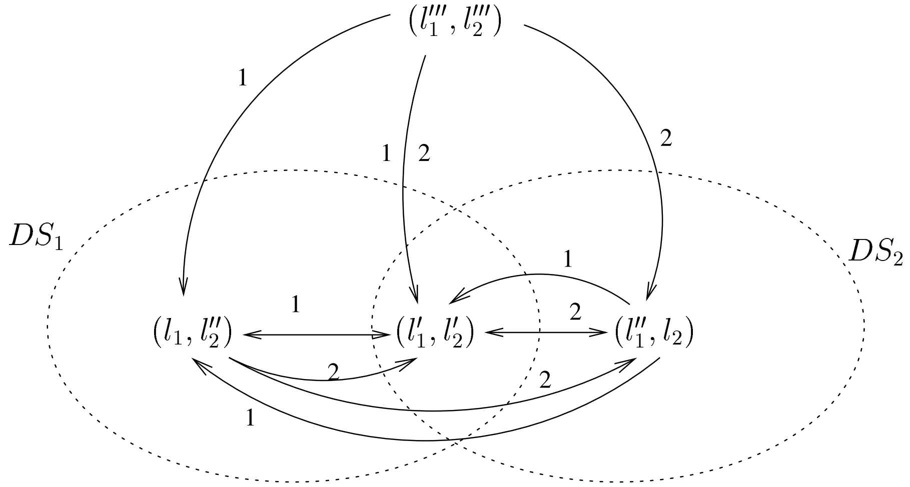

Figure 1: An example of deontic system and its generated frame. In the example above the environment is not considered and the local states for the agents are composed as follows. Agent 1: $L_1 = \{l_1, l_1', l_1'', l_1'''\}, G_1 = \{l_1, l_1'\}$. Agent 2: $L_2 = \{l_2, l_2', l_2'', l_2'''\}, G_2 = \{l_2,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

defining a family of modal operators representing a 'bird's eye view' of the knowledge of the agents is $S5_n$.

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 which picks out the global states in which agent i is running according 'correct (or acceptable) operating circumstances'. We illustrate this in Figure 1.

We can make use of the construction above to give an interpretation to the formulas of a deontic language in a way similar to what is done for knowledge on interpreted systems. 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 the truth of formula $\mathcal{O}_i \varphi$ at a global state signifies the truth of formula φ in all i related worlds, i.e., in all the points resulting from global states in which agent i is in a correct local state, i.e. in a green state.

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:

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

Validity on deontic systems is similarly defined on the class of generated frames.

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 systems

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.*

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, it turns out that the logic determined by deontic systems of global states is in fact stronger than KD45$_n$. Axiomatising this semantical class is a rather laborious exercise; we only report the main results here and refer the reader to [LS00] for further details.

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 $R(w)$ (where $R(w) = \{w' \in W \mid wRw'\}$).

(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 can be noted 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 2 *A frame F is i-j Euclidean if and only if $F \models 5^{i\text{-}j}$.*

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

Theorem 1 *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.*

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

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 3 *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 1, there exists a serial, secondarily universal model $M = (F, \pi)$ such that $M \not\models_w \varphi$, for some $w \in W$. By Theorem 2 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 p-morphism considerations, 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. ▫

We now turn to motivate the adequacy of the axioms of KD45i-j_n. 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, 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 can be checked that the logic KD$45_n^{i\text{-}j}$ contains also the following generalisation of axiom 4:

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

and indeed 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 KD$45_n^{i\text{-}j}$.

It is both instructive and useful to consider also what is likely to be 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 $IDS = (DS, \pi)$, 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 $\Box(\mathbf{g}_i \rightarrow \varphi)$. Each operator $\mathcal{O}_i$ can thus be defined as an abbreviation in terms of $\Box$ and $\mathbf{g}_i$ as follows:

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

$\mathcal{P}_i\, \varphi$ 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 $5^{i\text{-}j}$.

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

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

It would be reasonable to suppose that the S5 axioms for $\Box$ together with axioms Def.$\mathcal{O}_i$, D($\mathbf{g}_i$) and $\Box$- $\mathcal{O}_i$ provide a complete characterisation of the logic of interpreted deontic systems. We have not checked that this is so.

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 $X = A$.

4 Epistemic deontic systems

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, consider the usual clause:

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

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 S5$_n$ for the K_i component union (in the technical sense of [KW91, Gab98, GS98]) the logic 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:

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

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. 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.

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:

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

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

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

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 notions modelled in epistemic logic, positive and negative introspection for $\widehat{\mathrm{K}}_i^j$ do seem reasonable. Intuitively it is reasonable that "knowledge under the assumption of correct behaviour" should not imply 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 the previous pages.

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

$\mathrm{K}_i\,\mathcal{O}_j$ 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):

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

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:

$$\begin{array}{ll} \mathcal{O}_j\, \mathrm{K}_i\, p \rightarrow \widehat{\mathrm{K}}_i^j\, \mathrm{K}_i\, p & \\ \mathcal{O}_j\, \mathrm{K}_i\, p \rightarrow \widehat{\mathrm{K}}_k^j\, \mathrm{K}_i\, p & \quad (\text{any } k \in A) \end{array}$$

It is worth noting that, if we would like to give a complete characterisation in terms of Kripke frames of a language including $\mathcal{O}_i, \mathrm{K}_i$, and the modal operator $\widehat{\mathrm{K}}_i^j$, then the doubly-indexed operator 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, HM92b]. 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 [HM92b] 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 [HM92b], and it is reasonable to expect to have a logic whose fragments are KD45 for the $\mathcal{O}_i$ component, S5 for the K_i component, K45 for the $\widehat{\mathrm{K}}_i^j$ component and the interaction axiom (1).

5 Conclusions

In this paper we have tried to argue that interpreted systems are not only suitable models for representing knowledge, belief, and time, but can be extended to talk about the issue of compliance/violation, or in terms more common in computing, about the issue of correct functioning behaviour with respect to a protocol. We have explored the axiomatisation problem, and attempted to incorporate knowledge. Thanks to the semantical framework, this can be done not only in the straightforward way by means of the union of the two logics, but also by defining an operator on the intersection of the resulting two relations.

Although we have given the green/red labels a normative (evaluative) reading, they could also be read as normal/exceptional, and 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. This seems to be a concept worth further exploration, especially with respect to defeasible knowledge. This remains to be investigated together with the issue of completeness for the extension to the $\widehat{\mathrm{K}}_i^j$ logic.

Quite apart from this, many potential avenues for further work seem to be open. One is an exploration of the protocols (e.g., environments, or contexts) that guarantee particular knowledge states. Another is the definition of correct behaviour on runs rather than on states. We also believe it may be fruitful to examine the possibility of adding additional structure to the environment; in particular, we have in mind the possibility of analysing communication mechanisms. We should also like to understand better the relationship between obtaining the determination results directly and obtaining them via the reduction method discussed above.

On a more general level, we see there may be potentially two points of value in this contribution. First, there may be some mileage in bringing together deontic logic with a computationally grounded formalism such as the one of interpreted systems; indeed, the lack of a clear computationally grounded semantics is perhaps one of the reasons for which deontic logic has not found a greater role in computing. Secondly, the whole issue of the different subtle ways in which permission and knowledge can be combined seems fruitful. In this paper we began exploring some of these, but we believe others, perhaps equally compelling, still remain to be analysed.

References

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

[Aum76] R. J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239, 1976.

[BdR97] P. Blackburn and M. de Rijke. Why combine logics? *Studia Logica*, 59:5–27, 1997. Edited by D. Gabbay and F. Pirri.

[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.

[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.

[Hin62] J. Hintikka. *Knowledge and Belief, an introduction to the logic of the two notions*. Cornell University Press, Ithaca (NY) and London, 1962.

[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.

[HM92a] J. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[HM92b] 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.

[HMV97] J. Halpern, R. van der Meyden, and M. Y. Vardi. Complete axiomatisations for reasoning about knowledge and time. Submitted, 1997.

[Hoe93] W. van der Hoek. Sytems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195, April 1993.

[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.

[LS00] A. Lomuscio and M. Sergot. Investigations in grounded semantics for multi-agent systems specifications via deontic logic. Technical report, Imperial College, London, UK, 2000.

[Mey96] R. van der Meyden. Knowledge based programs: On the complexity of perfect recall in finite environments. In Y. Shoham, editor, *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 31–50, San Francisco, March17–20 1996. Morgan Kaufmann Publishers.

[MH95] 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.

[MW93] J.-J. Ch. Meyer and R. J. Wieringa, editors. *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons, 1993.

[Pop94] S. Popkorn. *First Steps in Modal Logic*. Cambridge University Press: Cambridge, England, 1994.