Deontic and action logics for collective agency and roles

José Carmo and Olga Pacheco
Dep. of Mathematics
Instituto Superior Técnico
Jose.Carmo@math.ist.utl.pt

and

Dep. of Informatics
Univ. of Minho
omp@di.uminho.pt

Abstract
In this paper we address the problem of collective agency, and propose a deontic/action modal logic for that purpose. We argue that once we want to attribute obligations (permissions or other deontic notions) to a set of agents, we need to consider a new agent – an institutionalized agent, and specify how he interacts with the external world: how the obligations flow from the institutionalized agent to the real agents that support him, and how the actions of the latter count as actions of the former. But an agent may act in many qualities (roles), and it is essential to know in which quality an agent has acted, or intends to act, for three main reasons: to know the effects of the act, its deontic qualification, and authentication issues. Thus, we extend the “sees to it” action operator with an explicit index that states the quality (role) in which the agent has acted. We also show how to associate obligations to roles, and illustrate how this can be used to express the desired flow of obligations.

Section 1. Introduction
The problem of collective agency is an important problem when we want to formalize organizations, and agents and organizations interaction. To our knowledge, the modal logic approaches that have been proposed to capture the notion of collective agency, and to model deontic notions (obligations, permissions, etc.) applied to sets of agents, have not been well succeeded. In our opinion, the main reason is the non recognition that only agent’s acting can be deontically qualified. If we want to attribute, for instance, obligations to a set of agents, seen as a whole, then we should consider a new agent – that we call institutionalized agent – and attribute to him such obligations. And Law seems to give us reason about this.

Of course, once we make that step, then we should stress how an institutionalized agent interacts with the external world: who represents and acts for him. The flow of the obligations from the institutionalized agent to the real agents that support him must be specified, as well as we must specify how the actions of the latter count as actions of the
former. But, and this is important, not all actions made by an agent, that supports an institutionalized agent, count as actions of the latter: the agent must have acted on the quality of representative of the institutionalized agent. The effects of the actions depend on the quality in which the agent has acted (on the role played by the agent), the deontic qualification of the act may also depend on the quality in which the agent has acted (the agent may be permitted to bring about some state of affairs playing some role, but forbidden when playing other roles), and, for that reason, that quality must be authenticated. For these reasons, we propose to extend the action modal operators of the “sees to it” type with an explicit index stating the quality (role) in which the agent has acted. We also show how to associate obligations to roles in our formalism, and illustrate how this can be used to express the desired flow of obligations from the institutionalized agent to the real agents that support him.

We should stress that we do not intend to provide a model for all forms of collective agency. Collective agency is a very complex subject with multiple problems. We will focus on the particular case of institutionalized (or, at least, organized) collective agency. On the other hand, although the importance of the notion of "acting in a role” is here motivated, in first place, within the context of the relationships between the acts of the agents that support an institutionalized agent and the acts of the latter, the proposed solution is more general, and is also useful for situations not involving agent’s interaction, as we shall illustrate.

The rest of this paper is organized as follows. In Section 2 we present the action operators (proposed by Kanger, Pörn and Lindahl, among others) that we want to use and extend. In Section 3 we discuss possible ways of defining the attribution of obligations, permissions and prohibitions to sets of agents, and show that neither of them is satisfactory. In Sections 4 and 5 we motivate the need of the notion of institutionalized agent, and of the proposed extension of the action operators, as well as we make a brief comparison with some other works, namely with works that use the concept of role. In section 6 we present our formalism and show how to associate the deontic operators to roles. Conclusions and directions for further research appear in Section 7.

Section 2. Action logic

In their pioneering work, Kanger, Pörn and Lindahl have combined deontic and action logics as basic building blocks to describe social interaction and complex normative concepts (see e.g. [Kanger 57, 72; Pörn 70, 77; Lindahl 77]). Their logics have sufficient expressive
power to be able to articulate several distinctions at an appropriate abstract level, mainly in
virtue of the modal logic of action they employ. They introduce a relativized modal operator,
here designated by $E_i$, where expressions of the form $E_i B$ are read “the agent $i$ brings it
about that $B$” or “agent $i$ sees to it that $B$ is the case”. The formal-logical development of this
approach is due to Pörn [Pörn 70], with more refined versions presented in [Kanger 72;
Pörn 77; Elgesem 93].

An important feature of these action logics is that actions are taken to be relationships
between agents and the states of affairs that they bring about. The “sees to it” modal operator
relates an agent with the effects of his action, omitting details about the specific action that
was performed (and setting aside temporal aspects).

This approach to the logic of action offers an expressive power rather different from that of
dynamic logic (see e.g. [Harel 79, 84]). For instance, using $E_i$ one can express several
different positions in which an agent $i$ might be with respect to a certain state of affairs $B$,
such as $E_i B$ (did), $E_i \neg B$ (averted) and $\neg E_i B \land \neg E_i \neg B$ (remained passive), as well as
notions of control of other agents, like $E_i E_k B$ (made $k$ do), $E_i \neg E_k B$ (made $k$ avoid), etc.
Moreover, combining $E_i$ with deontic operators we can then talk about the different
normative positions in which one or more agents might be, and use that to express legal
concepts and relations like rights, duties, etc., as has been done e.g. in [Lindahl 77].

For the definition of the logic of the operator $E_i$, different modal approaches have been
taken, providing in general non normal modalities ([Chellas 69] constitutes the main
exception to this). Briefly, Kanger and Pörn [Kanger 72; Pörn 77] define $E_i$ as boolean
combinations of two normal modalities; in the “stit theory” of Nuel Belnap and Michael
Perloff (see e.g. [Belnap 89; Belnap & Perloff 89]), the operator $E_i$ is primitive (denoted by
STIT$_i$) and semantically defined using tree-like models, including choice sets; tree-like
models are also used in [Chellas 69], in Chellas prior account to the modal logic of agency
(see also [Chellas 92]); and in [Elgesem 93] and [Santos & Carmo 96] the operator $E_i$ is
also primitive, but its semantics is defined using variants of the minimal models popularized
by [Chellas 80].

---

1 For a further brief comparison between these two approaches to the logic of action, see, for instance, [Brown & Carmo 96].
2 With respect to the theory of normative positions, and its importance, see also [Jones & Sergot 93] and [Sergot 99].
3 For an overview of action logics in this tradition see [Santos & Carmo 96].
As an illustration, in [Santos & Carmo 96] the models take the form $M = \langle W, f_1, \ldots, f_n, V \rangle$, where: $W$ is a non-empty set of possible worlds; $V$ is a function applying each propositional variable (or atom) $p_i$ to the set of elements of $W$ where $p_i$ is true; and each $f_i$, $i=1, \ldots, n$ (one for each agent), is a function $f_i : 2^W \to 2^W$, with $f_i (Z)$ (for $Z \subseteq W$) intuitively denoting the set of worlds where agent $i$ sees to it proposition $Z$ (these functions are then constrained in order to get the desired principles for the $E_i$ operator). The truth of an action formula $E_i B$ at a world $w$ in such model $M$ is then defined as follows:

$$M |=_w E_i B \quad \text{iff} \quad w \in f_i (\| B \|)$$

(there $\| B \| = \{ w : M |=_w B \}$). Similar models will be considered here, whenever we want to refer to the semantics of this kind of modal operators.

Although the formal properties assigned to the action operator $E_i$ vary among the different authors, all the different proposed logical systems have in common the following two axiom schemas:

\[ \begin{align*}
(T) & \quad E_i B \rightarrow B \\
(C) & \quad (E_i B \land E_i C) \rightarrow E_i (B \land C)
\end{align*} \]

and the (proof) rule:

\[ \text{(RE) If } |- B \leftrightarrow C \text{ then } |- E_i B \leftrightarrow E_i C \]

besides incorporating the tautologies and the Modus Ponens inference rule. Thus, we can see these principles as the core of any action logic (of this type). Moreover, most of the logics also include the schema:

\[ \text{(No) } \neg E_i T \quad \text{(where } T \text{ denotes a tautology)} \]

in order to express that the truth of $E_i B$ must imply that the action of agent $i$ was necessary to get the state of affairs $B$ (see [Elgesem 93] for a discussion on this). Since we are in

---

4 Following some other authors, we are here informally using the term proposition to refer to a set of worlds.
5 The (T) schema captures the intuition that if agent $i$ brings it about that $B$, then $B$ is indeed the case; that is, $E_i$ is a "success" operator.
accordance with this principle, we will call of \textit{minimal action logic}, a classical modal logic of type ECNo (using [Chellas 80] classification).

Some extensions and refinements of these action logics have been proposed \textsuperscript{6}. A natural extension of these logics, considered e.g. in [Lindahl 77], and that is relevant to our interest here, consists in allowing the indexing of the operator $\mathbb{E}$ by a (finite) set of agents, instead of only by a single agent. Informally, $\mathbb{E}_X B$ means that the set of agents described in $X$ jointly see to it the state of affairs $B$. In general, when we assert $\mathbb{E}_X B$ we want to express that the actions of the agents in $X$ cause the state of affairs $B$; the actions of each of such agents were necessary (or, at least, contributed in a significant way) to the production of $B$. We may say that the agents described in $X$ jointly cooperate to bring about that $B$ is the case \textsuperscript{7}.

Within such extension, we can express some notions of \textit{collective agency}, and define logics where formulas of the form

$$E\{i, j\} B \land \neg E_i B \land \neg E_j B \quad \text{(with } i \neq j\text{)}$$

can be consistent, allowing to express situations where two (or more) agents were able to jointly do some task (e.g. to move a very heavy table), without being the case that any of them has done it by himself \textsuperscript{8}.

\textsuperscript{6} For instance, [Santos & Carmo 96] propose a distinction between a direct action operator and an indirect action operator (the logics here considered are assumed to be neutral with respect to this), and [Santos, Jones & Carmo 97] propose a non-necessarily successful action operator $H_i$, with the following informal meaning: $H_i B$ means that agent $i$ attempts to bring about that $B$.

\textsuperscript{7} We are aware of the fact that joint action is a very complex subject and that there are multiple forms of \textit{jointly seeing to it}. In general, when we use the sees to it operator, indexed by a set of agents, to describe collective action, we abstract from a lot of things, and, in particular, we abstract from intentional issues. So, when we say above that "the agents described in $X$ jointly cooperate to bring about that $B$ is the case" that does not mean necessarily that such cooperation was intended by the agents in $X$ (see the next two footnotes for examples). We note, however, that not all authors agree with this position; for instance, Tuomela argues that the sees to it operator should be used only to describe intentional agency and intentional actions. Independently of the position we take regarding this general issue, we should stress that within the models of collective agency discussed in this paper, our use of action operators presupposes in general intention. As a matter of fact, in this paper we will be concerned in modeling organized collective agency, where the intended (or orchestrated) ways of acting associated to an organization, and that may vary from organization to organization, are predefined in the moment of its creation, and are expressed mainly by rules that state how actions of agents acting in some roles count as actions of the organization. And, since we assume that for an agent to act in a role he must show that he has the desired qualification (acting in the role of \textit{himself} constitutes an exception to this), acting in a role presupposes necessarily some kind of intention. Nevertheless, we do not want to go deeper on this problem here.

\textsuperscript{8} Another situation that could be described by a formula like the previous one, would be a typical criminal law "school case" like the following: it is known that only 100g (or more) of some poison is enough to kill a person; two agents $i$ and $j$ both put 50g of poison in a glass of water that mister $k$ was expected to drink; mister $k$ drank the glass with the poison and died. Such a situation would be correctly described by a formula like $E\{i, j\} \text{dead}(k) \land \neg E_i \text{dead}(k) \land \neg E_j \text{dead}(k)$, independently of the agents $i$ and $j$ were...
Naturally, formulas of the form
\[ E\{i,j\}B \land E_iB \quad (\text{with } i \neq j) \]
can be also consistent\(^9\), and there may be even cases where the production of some state of affairs \(B\) by an agent \(i\) “counts as”\(^10\) if the set of agents \(\{i, j\}\) has produced \(B\). However, by obvious reasons, we reject a general principle of the form
\[ E_XB \to E_ZB, \text{ for } X \subseteq Z \]

Using minimal models, of the type described above, it is trivial to give a semantics to such extension: just include in the models a function \(f_X\) for each finite set \(X\) of relevant agents.

**Section 3. Deontic logic**

The traditional approach to deontic logic sees it as a branch of modal logic, where the necessity operator is interpreted as meaning obligation, and denoted by \(\Diamond\). The dual of \(\Diamond\) (\(\neg\Diamond\neg\)) is then denoted by \(\Box\) and interpreted as meaning permission; and prohibition is expressed by an operator \(F\), defined as \(\Diamond\neg\). With respect to the logic of these operators, standard deontic logic (SDL for short) defines \(\Diamond\) as a normal KD modality. SDL gives raise to a set of paradoxes, and many other proposals have been made to try to solve them\(^11\). However, we do not want to enter in details on such issue, since it does not constitutes the topic of this paper (see e.g. [Hilpinen 71; Meyer & Wieringa 93; Carmo & Jones 98]).

Besides these impersonal deontic operators, we can also conceive similar personal deontic operators, indexing them with an agent, with the following informal meaning:
\[ \Diamond_iB : i \text{ is under an obligation of producing (doing) } B \]

---

\(^9\) Consider, for instance, a situation, similar to the one described in the previous footnote, where: agent \(i\) puts 100g of poison in the glass of water of mister \(k\); agent \(j\) puts 50g of poison in the same glass of water; mister \(k\) drank the glass and died. A possible way of describing such situation would be to assert \(E\{i,j\}\text{dead}(k) \land E_i\text{dead}(k) \land \neg E_j\text{dead}(k)\).

\(^10\) For a discussion of a “count as” operator see [Jones & Sergot 96].

\(^11\) This does not mean that SDL cannot serve some purposes. For instance, it may be argued that SDL can be used if we are only interested in expressing what ideally is the case, and only want to be able to express that a violation has occurred, without pretending to express what to do when some violation occurs.
\[ P_i B : i \text{ is permitted to produce (do) } B \]
\[ F_i B : i \text{ is forbidden to produce (do) } B \]

A natural question to pose is if such operators need to be primitive, or if they can be defined as combinations (iterations) of impersonal deontic operators and action operators. The obvious candidates are:

\[ O_i B = O E_i B \]
\[ P_i B = P E_i B \]
\[ F_i B = F E_i B \]

In [Herrestad 96; Krogh 97] two main criticisms are made against this option. First we loose the interdefinability of the relativized deontic operators. As a matter of fact, we still get

\[ P_i B \leftrightarrow \neg F_i B \]

(as well as \( O_i B \rightarrow P_i B \), if \( O \) satisfies the (D) schema), but

\[ F_i B \leftrightarrow O_i \neg B \]
\[ P_i B \leftrightarrow \neg O_i \neg B \]

are no longer valid (note that \( O_i \neg B = OE_i \neg B \) is much stronger than \( O \rightarrow E_i B = F_i B \)). We do not think that this is a crucial problem.

The second criticism may be called the “problem of transmission of obligations”, and may be described as follows. If we assume a normal modality for \( O \) (as in SDL), then the following schema becomes a theorem (according to the abbreviations above):

\[ O_i E_j B \rightarrow O_j B \]

which is clearly unacceptable. However we note that this problem only appears if we assume that \( O \) satisfies the (RM)-(proof)rule (i.e. that \( O \) is closed under implication, in the sense that \( \models \neg B \rightarrow C \) implies \( \models \neg O B \rightarrow O C \)). We can define non-normal logics for \( O \) where this is not the case (even keeping, if we wish, that \( O \) satisfies some weaker versions of the (RM)-rule), as well as we may be only interested in applications where we do not want to have action
operators in the scope of other action operators (for instance, because we are not interested in notions of control).

Even if we do not have a definite position regarding this issue, in this paper we want to explore, and to extend, this possibility of defining personal deontic operators by iterating impersonal deontic operators and action operators of the “sees to it” type.

The following step, since we are here interested on the topic of collective agency, is to attach deontic operators to a (finite) set \( X \) of agents, and try to define them using the previous deontic and action operators. Consider, for instance, the obligation operator \( O_X \).

What is the meaning of \( O_X B \)? A first hypothesis is to define this kind of collective obligation in terms of individual obligations of the members of \( X \). Two options are then possible (where the for all operator below may be seen as an abbreviation of an iterated \( \land \), assuming \( X \) finite):

\[
\begin{align*}
i) & \quad O_X B = \forall x \in X \ O x B \quad (= \forall x \in X \ O E x B) \\
ii) & \quad O_X B = \exists x \in X \ O x B
\end{align*}
\]

Consider first ii)\(^{12}\). Not only it validates \( O_X B \rightarrow O_Z B \), for \( X \subseteq Z \), as it is of no use in practice. What is the interest of knowing that one element of the set \( X \) has an obligation of producing \( B \), if we do not know whom he is. If the obligation is not fulfilled, who is the responsible? (And, if we think in terms of identical definitions for the collective permission and forbiddance operators, easily we get situations where \( P_X B \) and \( F_X B \) are both true\(^{13}\).)

Consider now i). Although in many cases we may see \( O_X B \) as an useful short way of expressing the same as \( \forall x \in X \ O x B \) (abbreviation that we will also consider, later, in this paper), it is hard to accept that this definition traduces the notion of collective obligation, in the sense of an obligation on a collective agency. And this becomes obvious when we think in terms of non-fulfillment, or violation, of an obligation of such kind. Consider, for instance, as a very simple example, a situation where a couch says to his football team: “You are obliged to mark (at least) five goals on today’s game”, obligation that we could express (using \( p_i \) for player \( i \), etc.) by: \( O \{ p_i : i=1,\ldots,11 \} \) five. Is it correct to infer that every player (or even any player) is under an obligation to mark five goals? And how we define

\(^{12}\) The “weak collective obligation” considered in [Royakkers and Dignum 98] gives raise to this definition.

\(^{13}\) But, of course, we could define, for instance, \( F_X \) through i) and \( P_X \) through ii).
violation of such an obligation? Surely that a situation where $E\{p_i: i=1,\ldots,11\} \text{five} \land \forall_{i=1,\ldots,11} \neg E\{p_i\text{five}\}$ is true, does not count as a violation of $O\{p_i: i=1,\ldots,11\} \text{five}$.

The previous example suggests that we follow an approach similar to the one taken above for the definition of $O_i B$, and define:

$$iii) \quad O X B = O E X B$$

(and similarly for the other deontic operators). However, this solution is not behind criticisms. Suppose that a violation of $O X B$ occurs (what we might express by $O E X B \land \neg E X B$). Who is the responsible of such a violation? What does it mean to simply say that it is the set $X$ that is responsible for the violation? Who is possibly subject to punishment, because a violation occurs? $X$, or some member of $X$, or all members of $X$? These questions do not seem to be essential in the example of the football team above, because in such simple example the problem of violation and punishment does not seem so relevant, but they may be crucial in many day-life applications. In some sense this problem suggests that we need to relate again the collective obligations with individual obligations of the members of the set.

Consider now prohibitions (and permissions). How to interpret situations like $F X B \land \forall_{x \in X} \neg F x B$?

Of course, we could say that collective deontic operators, like $O X$ and $F X$, need to be primitive, and cannot be defined using the other deontic and action operators. But, clearly, that does not solve the questions we have just posed. At this moment, we may start to think that the problem relies not on the deontic component, but on the notion of collective agent seen as a set of agents. We come back to this issue in the next section.

Before closing this section on deontic operators, we would like to stress that obligations, permissions and prohibitions are far from exhausting the set of relevant deontic notions in which we might be interested. Another interesting deontic operator would be an authorization operator $A$, where $A B$ informally means that $B$ is “authorized” and $A_{i} B = A E_{i} B$ means that the production of $B$ by the agent $i$ is authorized. Of course, this interpretation must have some consequences, imposing that agent $i$ must have some “rights” regarding the
production of $B$, possibly meaning that the means for that must be provided if, and when, agent $i$ decides to see to it that $B$. Naturally, this suggests that there is a bi-direction here, and that we need to express in the authorization operator who has given such authorization to $i$, to whom $i$ has those rights. In [Henning 96; Krogh 97] a double index of the deontic operators is proposed precisely to express notions like rights, duties, etc. However, here we want to avoid such double index and to avoid entering on the discussion of such non trivial concepts (rights, duties, etc.), in order to focus on the (also non trivial) problem of collective agency. Thus we will not introduce any explicit authorization operator. We limit ourselves to informally say that a sentence $B$ is authorized (according to some deontic specification - seen as a set $\Gamma$ of sentences of our deontic/action formal language) iff $B$ is authorized in the weaker sense that we can derive explicitly from $\Gamma$ that $B$ is permitted (and not only that we cannot derive that $B$ is forbidden), that is, iff $^{14} \Gamma \models \mathcal{P} B$. In general, we will be interested in applying this notion of authorization with respect to action sentences expressing that an agent $i$ sees to it some state of affairs acting in some quality/role (to be introduced later).

Section 4. Collective agency and collective agents

The following quotation from [Bailhache 91, page 80] (taken from [Royakkers and Dignum 98]) illustrates what we think it is the usual wrong direction regarding collective agency: “For imagine, for example, that such masons ought to build the foundation of the house, such roofers to assemble its roof, such electricians to set its electrical network, and so on: the firm, as the set of all these workers will have the duty of building the whole house. ...”

*It is precisely this identification of the firm with the set of his workers that we think it is the cause of most of the problems*, and of the wrong direction generally taken regarding collective agency. And, just as a first argument, we note that the set of workers of the firm may change, without changing the firm. There exist here two related, but distinct, entities: one a set of human agents – the workers - and another an institutionalized agent – the firm.

We can have situations where a set of agents jointly acting produce some state of affairs, as (for instance) when two agents change from place a very heavy table. The attachment of a set of agents to the sees to it operator $E$ allows us to express situations like this $^{15}$. But this

---

$^{14}$ Where $\Gamma \models \neg C$ iff there exists a finite number of hypotheses in $\Gamma$, $B_1, \ldots, B_n$, such that $\neg B_1 \land \ldots \land B_n \rightarrow C$ (or, in another way, if $C$ can be derived using only the theorems, the hypotheses in $\Gamma$, and the Modus Ponens inference rule).

$^{15}$ Maybe the situation where two agents produce a contract should be modeled in this way, but we do not have a definitive position about this.
situation is distinct from the cases, in which we are mainly interested here, where we want to look to a set of agents as acting as a single (collective) entity, and to whom we want to associate obligations (permissions, etc.), as in the case of the firm. In our opinion, deontic notions like obligations, prohibitions, responsibilities, authorizations, etc. are only meaningful when associated to agents. As soon as we want to assign obligations (or other deontic qualification) to a set of agents, what we need to do is to consider a new entity – a new agent – associate to him the desired obligations, and specify how he interacts with the external world, specifying who represents him and to what extent (who has the power to act on his name) and how his obligations become obligations of the “real” agents that supports him (i.e. that acts for him).

In order to support our intuitions we decided to look at Law with respect to these matters. Because Law tries to “model” the real world we knew that collective agency could not be avoided and that some legal models of collective agency would certainly exist. The main idea was to look in the legal domain for models and methodologies of collective agency, to capture its main aspects and try to incorporate them in our models and formalisms.

From our digression on Law, we concluded [Pacheco & Carmo 98] that, in Law, in addition to "natural persons" there exist "artificial persons", which are entities that aggregate several persons allowing them to collectively pursue some interests. An artificial person is a person having juridical personality and legal competence as any natural person. By default, an artificial person has a typified structure, formed by a set of organs and a set of norms regulating its behavior. This abstract structure is supported by real persons. An artificial person acts through other agents: the “titulars” of the organs of its structure (the holders of such positions). The norms governing the relationships between the artificial person and the titulars of its organs are predefined and must be known by the others agents of the society.

So, according to Law, we should distinguish between natural persons and artificial persons – that we prefer to call here of non-institutionalized and institutionalized agents. In a paper in preparation [Pacheco & Carmo 99] we intend to discuss in more detail the differences between the two, and to set a kind of specification language for agents that covers both the typical structure of an institutionalized agent, and some typified forms of interaction (and namely the legal relationships of mandate and representation that specifies the forms of interaction between the institutionalized agent and the agents that support him – see below). Here we want to concentrate on what is essential to our discussion related to deontic and action logic, and two characteristics taken from Law seem to be fundamental: 1) an
institutionalized agent may be the subject of obligations and responsible by their non-fulfillment/violation; and 2) an institutionalized agent always acts through other agents (typically the titulars of its organs).

Since the fulfillment of most obligations presuppose the performance of some actions, and since an institutionalized agent never acts directly, we need to have some mechanism that stresses how the obligations flow from the institutionalized agent to the titulars of its organs – and Law provides a general figure of interaction between agents that also applies in such situation: the mandate relationship. Space limitations prevent us of entering in details about this relationship (see [Pacheco & Carmo 99]). Nevertheless, at the end of this paper we will show how we can express in a simple and elegant way such flow of obligations, by using an appropriate notion of role.

On the other hand (as another direction of the problem), since an institutionalized agent always acts through other agents, we need to have some mechanism that stresses how the actions performed by these count as actions performed by the institutionalized agent. That is, we must clarify the way actions of the titulars of its organs affect the institutionalized agent: do not forget that it is the institutionalized agent that is responsible by the non-fulfilment/violation of his obligations, even if it was not him that has directly performed (or not performed) the relevant actions. Again, Law provides a general figure of interaction between agents that also applies in such situation: the representation relationship. The titulars act as representatives of the institutionalized agent, which briefly means that they have powers to act on behalf of the institutionalized agent.

The question is if, and how, we can model in our logical language this flow from the actions of the titulars of the organs (for instance, the administrators) to actions of the institutionalized agent. Assuming that \(i\) denotes an agent (institutionalized or not) and \(k\) an institutionalized agent, and that we have a predicate \(\text{is-administrator-of}\) (and that \(B\) describes a state of affairs within the scope of the administration tasks – we will refer later to this issue), we could try to express that by a formula of the form:

\[\neg \text{DE}_i B, \text{for any institutionalized agent } i \text{ (and any formula } B)\]

---

16 If we introduce a direct action operator \(\text{DE}_i\) (adapting the ideas in [Santos & Carmo 96]), then we could logically capture this idea that an institutionalized agent never acts directly, by imposing the following axiom schema: \(\neg \text{DE}_i B\), for any institutionalized agent \(i\) (and any formula \(B\)).

17 Briefly, the mandate is a legal relationship between two persons by which one person is obliged to perform a particular set of acts on the account of the other.

18 There is no reason why an institutionalized agent cannot be a titular of an organ (for instance, the administration) of another institutionalized agent.
(*) \( \text{is-administrator-of}(i,k) \rightarrow (E_i B \rightarrow E_k B) \)

As a first remark about the previous representation, we note that we are aware that probably we need to replace the second material implication, \( \rightarrow \), by a “count as” operator of the kind proposed in [Jones & Sergot 96], \( \Rightarrow \), and write instead:

(**) \( \text{is-administrator-of}(i,k) \rightarrow (E_i B \Rightarrow E_k B) \)

However, for simplicity reasons, we do not consider here such “count as” operator, since we think that the main issue that we want to address in this paper is not concerned with the use of \( \rightarrow \) or \( \Rightarrow \).

The main reason why (*) (or (**)) fails to represent what is pretended, is the fact that the action operator used does not stress the quality (the role) in which \( i \) has acted when bringing about \( B \). The sentence \( \text{is-administrator-of}(i,k) \) expresses that \( i \) possesses the quality of being administrator of \( k \) (\( i \) has that qualification), and so \( i \) can play the role of administrator of \( k \), but that does not mean that “\( i \) has acted on that quality (on that role)”. Agent \( i \) may possess many qualities and produce a similar act acting in different qualities. But the quality in which \( i \) has acted (or intends to act) is fundamental with many respects (discussed below).

Let us then suppose that we extend our action operators by explicitly indicating in them the quality used by the agent (the role played by the agent) to bring about the referred state of affairs, allowing to express formulas like (a more concise syntax for these action formulas will be presented later):

\[ E_i \text{ acting as administrator-of}(k) B \]

As general principles (that will apply with respect to any quality, and not only to the quality of being administrator), we assume:

i) \( E_i \text{ acting as administrator-of}(k) B \rightarrow E_i B \)

(If \( i \) sees to it that \( B \), acting in some quality, \( i \) sees to it that \( B \) \footnote{We can see \( E_i B \) as meaning that \( i \) brings about that \( B \) acting in some quality that he posses. It is arguable if, within the scope of a "logic of acting in a role", it is meaningful to also consider an operator like \( E_i \). For the moment we keep it, together with this general bridging principle i), but it might be the case that this introduction of two kinds of sees to it operators has no real advantages.})
ii) \( E_i \text{ acting as administrator-of}(k) \Rightarrow is\text{-administrator-of}(i,k) \)

(If \( i \) sees to it that \( B \), acting in some quality, \( i \) possesses that quality)

As we have referred, the identification of the quality in which \( i \) has acted is important with many respects. In what follows we mention just three, that we think are fundamental.

1) Effects (and juridical consequences) of the act:

The effects of an action performed by \( i \) will depend on the quality \( i \) has acted. Thus, although (*) (or (**) should not be valid, it is correct to say that an action performed by \( i \), as administrator of \( k \), counts as an action performed by \( k \), which can be expressed by

\[(***) \ E_i \text{ acting as administrator-of}(k) \Rightarrow E_k B \]

Then, supposing that \( E_i \text{ acting as administrator-of}(k) \) is the case, we can deduce (from i) and (***) both \( E_i B \) and \( E_k B \). But this is not yet what we want. First, it does not allow us to discriminate \( i \) and \( k \), with respect to the production of \( B \), in order to analyze the juridical consequences of the act. Second, since we are trying to define a logic for the concept of acting in a quality (role), where we may assume that when an agent acts, he always acts in some role, we would like to know in which role we should say that \( k \) has brought about \( B \), in virtue of the fact that one of his administrators has brought about \( B \) (in that quality).

---

20 One of the referees suggested that we should have some kind of dual property of (***) like "if \( k \) is an institutionalized agent and \( k \) has brought about \( B \), then there exists some agent \( i \), acting as representative of \( k \), such that \( i \) has brought about \( B \)". If we want to express such property in our formal language, to be presented later, we would be faced with two problems. One has to do with the fact that there can be various qualities (roles) that can be considered as special cases of a representative relationship (role), like "being administrator of", "being the president of", etc. Thus, in order to describe that property succinctly we would need to extend our language with a notion of subrole (subquality), or a similar notion, and to establish the appropriate relationships between acting in a role and in a subrole - we should impose something like "if an agent \( i \) acting in a subrole (a more specific role) brings it about that \( B \), and if it is permitted to bring about \( B \) simply acting in a more general role, then it counts as if agent \( i \) has brought about \( B \) acting in the more general role", as well as "if it is permitted to bring about that \( B \) acting in some role, then it is also permitted to bring about that \( B \) acting in a more specific role (a subrole)". (Note also that for a role \( r_1 \) be a subrole of a role \( r_2 \) it is not enough that any agent that posses the quality corresponding to \( r_1 \) also possess the quality corresponding to \( r_2 \) - material implication between such qualifications is not enough - otherwise any role would be a subrole of the role himself - introduced below, which we do not want). This is a natural and interesting extension of the logic here proposed, and is actually under research. A second problem is that, since we have the schema (C) for the "sees to it" operator, if \( B \) is a conjunction, then it might have been the case that different representatives of the institutionalized agent have brought about different "parts" of \( B \). Another possibility to express similar properties would be to explore an extension of our logic with a direct action operator (see a previous footnote).
In fact, what we want to express is that when \(i\) acts as administrator of \(k\) (more generally, when \(i\) acts as representative of \(k\)) that counts as if it was \(k\), himself, that was acting. By introducing a special role himself, we can then state naturally\(^{21}\):

\[
\begin{align*}
\text{iii)} & \quad E_i \text{ acting as administrator-of}(k) \implies E_k \text{ acting as himself } B \\
\text{iv)} & \quad E_i \text{ acting as administrator-of}(k) \implies \neg E_i \text{ acting as himself } B
\end{align*}
\]

(Note that the previous principles, associated to acting as administrator-of, cannot be generalized directly to acting in others qualities\(^{22}\). Note also that, from iii) and from the generalization of i) to any quality, we can deduce (***).)

Although we are not proposing here any operator to express such notion of “juridical consequence” (which constitutes topic for interesting further research), we may assume intuitively that the effects of an act affect mainly the agents that act in their own name, which we can now capture. (Note that when \(i\) acts as representative of \(k\), the juridical consequences of \(i\)’s action will be attached to \(k\), and not to \(i\)).

2) Deontic qualifications

An agent may be authorized or even obliged to bring about \(B\) acting in some qualities (playing some roles), but not authorized to bring about \(B\) in other qualities. And an agent \(i\), that was not authorized to bring about \(B\), by becoming administrator (or president) of \(k\), may become authorized to bring about \(B\) when acting on that role (for instance, we may suppose, because \(k\) was authorized to bring about \(B\)).

3) Authentication

Since an agent may be authorized to bring about some state of affairs acting in some qualities, but not in others, it becomes essential to determine the quality in which he intends to act, and that take us to the authentication issue. This is particularly important in

\(^{21}\) As we shall see at section 6, we will allow to write formulas like \(E_k \text{ acting as } k B\) but (always) as an abbreviation of \(E_k \text{ acting as himself } B\) (in the formal language introduced at section 6 “acting as” is also replaced by “:” in order to get a more concise syntax).

\(^{22}\) With respect to principle iv), for some roles \(r(\ldots)\), it may be possible (and useful) to accept that

\[
E_i \text{ acting as } r(\ldots) B \land E_i \text{ acting as himself } B
\]

can be consistent. But, at least, we will reject the validity of a general principle of the form:

\[
E_i \text{ acting as } r(\ldots) B \implies E_i \text{ acting as himself } B.
\]
some computer applications, being object of important research (see e.g. [Abadi et al. 93, Lampson et al. 92, Massacci 97]), but also in everyday life examples without any relation with computers. By authentication we mean that for an agent to be able to bring about something, that he is authorized to do when acting in some quality, it is not enough that the agent possess that qualification: he may be required to show that he possesses it. What principle ii) traduces is the qualification expected to be authenticated when it tries to act in that quality. With the exception of the quality of is-himself, that every agent satisfies, we may assume that any other quality may need authentication.

Before proceeding, we note that in the same way that we can allow the “sees to it” operator \( E \) to be indexed by a set of agents, and not only by a single agent, we also can allow \( E \) to be indexed by a set of agents each acting in some quality. However, we should not allow \( E \) to be indexed by a set of agents, where such set is acting in some quality. One of the main reasons is precisely the referred authentication issue. For us, it makes no sense to require a set of agents to show that it possesses some qualification. Either that can be reduced to show that each member of the set has the desired qualification (and we are in a case of a set of agents each acting in some quality, that is the same for all the agents), or, if not, we are probably in presence of an institutionalized agent acting in some quality, which is simply a particular case of an agent acting in some quality, and can be expressed in our language. However, to simplify, in the rest of the paper we will not consider the operator \( E \) indexed by a set. We will focus on the collective agency problem in the context of institutionalized agents, where collective agency will be modeled in a different manner through the concept of "action in a role".

Section 5. Qualities, roles and acting in a quality (role)

The quality of being administrator (or titular of another organ) is only one of many qualities that may be relevant when describing agents acting and interacting, and the notion of "acting in a quality (role)" is relevant in many contexts, and not only within the context of institutionalized agents, although this was our first motivation in this paper.

In fact, we can distinguish two types of qualities with interest for describing agents acting and interacting: 1) qualities that express properties that an agent may have, independently of the others; and 2) qualities that express relationships between two \(^{23}\) agents. In both cases, those properties/relationships may depend on (include) other type of non-agentive

\(^{23}\) We can also think of relationships between more that two agents. But here we do not consider them.
information. And those relationships may relate any kind of agents, and not only a non-institutionalized agent and an institutionalized agent (e.g. a representation relationship, as any kind of contractual relationship, can be generically established between any two agents).

Important examples of 2) are typified legal relationships like mandatory-of, representative-of, etc., and we may wish to specify in our logic the general principles that govern them. But many other examples exist that are relevant in some applications, like worker-of, president-of, etc. As examples of 1) we have father, policeman, owner-of (we are here thinking in owner of a building, and not in a relationship between two agents), etc.

As further illustrations of the interest of the notion of acting in some quality (acting playing some role), we note, for instance, that the (legal) effects of having a car accident when acting as himself are completely different from the effects of having a car accident when acting as a worker of some firm. And for selling a house someone must act as owner of it (or as representative of the owner of it), since that quality need to be authenticated in order to validate the act.

The last example – “agent i acting as representative of an agent k that is owner of a building xpto, sells xpto” – may be rephrased by saying that – “agent i acting on behalf of agent k (i.e. representing k), when k acts in the quality of owner of xpto, sells xpto” – thus indicating that for the authentication of the act we must verify two things: the representation qualification of i and the ownership qualification of k.

Consider now a situation where i is representative (or titular) of an institutionalized agent k that is also related, suppose by a titular relationship (administration or whatsoever), with another institutionalized agent z. We may wish to be able to express (and to distinguish) two different situations:

a) i bringing about B acting as representative of k

b) i bringing about B acting as representative of k in his quality of (when k acts as) titular of z

This may be important for various reasons. First, a) may be authorized, and b) not. Second, when b) is the case, not only two qualifications need to be verified, as we want to be able to deduce (from the specific properties of the quality of “being representative of”) that it implies (count as):
c) k bringing about B acting as titular of z

and, from that, deduce (from the specific properties of the quality of “being titular of”) that it implies (count as):

   d) z bringing about B acting as himself

These examples suggest that we would like to generalize our notion of acting in some relationship with another agent, to a notion of acting in a relationship with an agent when he plays some role. Acting in some relationship with another agent could then be seen as acting in a relationship with an agent when he plays the role of himself.

Although we have not yet confirmed that Law supports this generalization, we think that it is useful, and in what follows we are going to consider it. (In the next section we will show how this generalization can be formally done.) Now we just refer that, using such generalization, we will represent a) and b) as follows (although with a more concise syntax):

   a) Ei acting as representative-of(k acting as himself) B
   b) Ei acting as representative-of(k acting as titular-of(z acting as himself)) B

The simpler formulas below can still be written, but they will be seen as abbreviations of the formulas above (see next section).

   a) Ei acting as representative-of(k) B
   b) Ei acting as representative-of(k acting as titular-of(z)) B

Before proceeding, we should also note that we are aware that in reality we never have the case where an agent i represents an agent k for everything. So, in general what we have is a qualification of the form is-representative-of-for(i,k,S), where S is a non-agentive sentence (and not simply a term, as we allow now, for other order of reasons 24).

Of course, when i brings about B, acting as representative of k for S, this must imply that B is in the scope of S. To formalize this we need a “scope” operator. We have some ideas about how this can be done, but we omit them from this paper, to simplify and also because it still needs further research. Nevertheless, we stress that in our logic, as it stands now, it is

24 For instance, to express that agent i is owner of building xpto: is-owner(i,xpto), or to express that agent i is the associate number n of the institutionalized agent k: is-associate-of(i,k,n).
possible to express that an agent \( i \) is representative of \( k \), without being permitted to see to it that \( B \) in that quality\(^{25}\), i.e. the schema:

\[
is\text{-representative-of}(i,k) \rightarrow P \text{E}_i \text{ acting as representative-of}(k) \ B
\]

is not necessarily valid. This may be seen as a way of solving that problem (even if only partially).

Finally, we would like to make some brief comments about related work. (A more detailed comparison between our approach and the works referred next, and others, will appear in [Pacheco and Carmo 99].)

Our model of institutionalized agents was inspired by the legal domain, in particular by the "artificial person" legal concept (see [Pacheco & Carmo 98] for a more detailed discussion). In general this model is in accordance with models discussed by Tuomela for the same type of collective entities ([Tuomela 95], e.g.): an institutional collective entity acting via its organization, having social positions to be filled by persons, characterized by social norms determining the tasks of the position-holders and their social roles and being represented by some position-holders (its representatives) that act on the behalf of the collective. Although the terminology is sometimes different, the general concepts are similar. In these institutionalized agents some of the complex problems discussed in collective agency (joint action, intentional action, commitment to act, ... for example) are simplified because they have predefined authority systems (group-will-formation systems - see [Tuomela 95] for a detailed discussion), which are known by all members of the collective entity and are also public to the others agents of the external society\(^{26}\). (We should stress that Tuomela's work also covers others forms of collective agency.)

Some authors, like [Rao et al. 92], also present the idea of a set of agents taken as a whole and as an agent at a different level. There are some differences between their proposal and ours. First of all, we do not reduce our collective (institutionalized) agent to the set of its

\(^{25}\) Note also that the fact that an agent \( i \) is not permitted to bring about \( B \), acting as representative of \( k \), is compatible with the fact that if, nevertheless, \( i \) succeeds in doing \( B \), in that quality, then it counts as an act of \( k \). (See [Jones & Sergot 96] for related issues.)

\(^{26}\) When an agent acts in a particular role of representative of an institutionalized agent he must "declare" that he is acting in that role, it is possibly to verify if he holds such position, and it is also possible to verify if he is acting in accordance with the "scope" predefined for the position he holds (if he is authorized to perform such acts using that position). There are predefined norms determining what is expected from an agent supporting a particular role and what to do if he does not act in accordance with them. When an agent accepts to hold a position he commits himself to accept those norms and to act in accordance with them.
individual elements: it has a different and intrinsic identity which does not change if we change the agents that hold the positions (roles) of the institutionalized agent. Another difference is that (in some sense) our institutionalized agent is not an agent of a different level, being an agent among other agents: he may interact with other agents at the same level (establish a contract with other agent, e.g.), hold a position in another institutionalized agent, be the subject of obligations, rights, or other normative concepts, and we do not have different logical operators to deal with institutionalized agents.  

With respect to roles, we should first stress that this term has been used by various authors, but not always with the same meaning and goal. For many authors (e.g. [Cuppens 94; Bertino et al 97]) role is a kind of “virtual agent”, and a role can bring about some state of affairs (the role is seen as acting on behalf of the agent that plays that role). Thus, for instance, with their formalism it would be possible to say that father (a role) can bring about some state of affairs. We think that this is intuitively unnatural to say. Moreover, to see “roles” as agents introduces a dangerous proliferation of agents. We know that we also introduce institutionalized agents, but these agents are not virtual: for all purposes, namely legal, they are agents like any others.  

Instead of saying that a role is a virtual agent, we could identify a role with a set of agents (the set of agents that may play that role – i.e. that satisfy the correspondent quality). Although we will be interested in refer to such set, we do not want to make such identification, because there may exist cases where a same set of agents may satisfy two different qualities, but for authorization and authentication purposes it is relevant to know in which of those qualities an agent of that set is acting (for instance, a same agent may be the owner of two different buildings, but when acting as owner of one he cannot sell the other). Our proposal is to identify a role that an agent can play with a quality that the agent posses (and can exercise), where such quality may correspond to a property that the agent satisfies, independently of the other agents, or to a relationship with other agent in which the agent is engaged (see next section for formal details).  

On the other hand, in some works [Skarmeas 95; Bertino et al 97] roles are seen as a way of structuring tasks and task decomposition, and other works [Skarmeas 95] use some kind of

---

27 But, of course, there are differences between an institutionalized and a non-institutionalized agent: the former cannot act directly; he only acts through his representatives, e.g..  
28 For instance, [Cuppens 94] see roles as virtual agents, and use them to give a semantics to the deontic operators where, for instance, $\text{PE} \downarrow B$ is true at a world $w$ if there is a world $w_1$ and an authorized role $r$, associated with $i$, such that $E_r \uparrow B$ is true at $w_1$, where it is assumed that, in $w_1$, $r$ is acting in place of $i$.  

---
roles, called social positions, that can be seen as a way to collect obligations (responsibilities) at an abstract level, which become then obligations of the agents that are assigned to that role. We are also interested in this latter application of roles. Tasks are different from roles, but it is natural to associate to a role tasks and capabilities: informally, the former can be seen as what the holders of the role are obliged to do (playing that role) and the latter what they are permitted to do (i.e., in some sense, the "scope" of the role). At the end of next section we will show how we can associate obligations and permissions to roles, within our formalism. Others common interpretations of tasks, in association with roles (where, informally, a task is seen as a procedure and roles are used to assign responsibilities for the execution of parts of that procedure), and task decomposition, are not of our concern here.

Finally, some works consider at the syntactic level a notion similar to bring it about acting in a role, and others (like [Cuppens 94]) not. For instance, in [Lampson et al. 92; Abadi et al. 93; Massacci 97] a logic for access control in distributed systems is defined, where an agent may say something on behalf of another agent, which can be seen as a notion similar to bring it about acting in the particular role of representative-of.

Section 6. The deontic/action language and logic

Our formalism can be defined as follows.

Formal language

The desired deontic/action language will be designated by \( L_{DA} \). (We will omit the description of its alphabet, since it can be deduced easily from what follows.)

The non-modal component of \( L_{DA} \) is a first-order many-sorted language [Enderton 72], \( L \), with a set of non-agentive sorts plus a distinct sort \( A_G \) (the agent sort). This is the language

---

More precisely, their "says" operator should be seen as an "attempt" operator. Besides many other differences between their approach and ours, we would like to refer that, according to them, when an agent acts playing some role, he restricts his privileges (and this has to do with the fact that, according to their approach, when an agent attempts to do something, without specifying in which role, that implies that he attempts to do it in all possible roles that he can play). Although in some situations this might be true, in general it isn't: when an agent \( i \) acts as administrator of a firm, he has different privileges (in some sense more).
where we express factual descriptions and properties and relationships between agents (institutionalized, or not: we do not distinguish them in what follows 30).

At least for the moment, we do not consider function symbols, having \( Ag \) as co-domain sort, in \( L \) (they do not seem essential to our purposes). For each sort \( s \) we assume an infinite number of variables (we write \( x^s \) to refer to a variable of sort \( s \)), and possibly some constants (constants of sort \( Ag - i, k, ... - \) are assumed to denote - are the names of - concrete agents, institutionalized or not).

We call the predicate symbols, involving only one parameter of sort \( Ag \), of agent property symbols, and those involving two parameters of sort \( Ag \), of agent relationship symbols, and call both of quality symbols (and we will generically denote them by \( q, q_1, ... \)). Basically as syntactic sugar, we assume that the quality symbols take the form “is-“ followed by some symbol (word 31), that we call of a role symbol or role generator (and generically denote it by \( r, r_1, ... \)). Moreover, we assume that the parameters, of sort \( Ag \), of these predicates, are their first parameters; that is, a quality symbol \( is-r \) will be of sort \((Ag, s_1, ..., sn)\), or of sort \((Ag, Ag, s_1, ..., sn)\), for \( n \geq 0 \) and \( s_j \neq Ag \), for \( j = 1, ..., n \). We also assume that we have one special agent property symbol: \( is-himself \), of sort \( (Ag) \).

The set of terms of each sort and the set of (well-formed) formulas of \( L \) is defined as it is usual (in the sequel, we use \( t^s, t_1^s, u^s, u_1^s, v^s, ... \) whenever we want to explicitly indicate that we are in presence of a term of sort \( s \), and use \( t, t_1, u, u_1, v, ... \) when we want to refer generically to a term of an appropriated sort). We call of qualifications the atomic formulas built from the quality symbols.

We then introduce two new sorts: \( R \) (role) and \( AgR \) (agent in a role, or agent playing a role, or agent acting in a quality). And, mainly to simplify the presentation below, we attach a sort to the role symbols as follows:

- if \( is-r \) is of sort \((Ag, s_1, ..., sn)\), then (we say that) \( r \) is of sort \((s_1, ..., sn \rightarrow R)\)
- if \( is-r \) is of sort \((Ag, Ag, s_1, ..., sn)\), then \( r \) is of sort \((AgR, s_1, ..., sn \rightarrow R)\)

30 But, naturally, this can be done, for instance, by introducing a predicate \( is-institutionalized \), of sort \((Ag)\), to that effect.
31 The names for the agent relationship predicates will terminate normally with “-of”, but this is not relevant for our purposes now. On the other hand, in the case of these predicates, the arguments, of sorts distinct from \( Ag \), are used mainly in situations where we want to express that agent \( i \) is the associate number \( n \) of the institutionalized agent \( k \), which may be represented as follows: \( is-associate-of(i, k, n) \). Thus, in general, the agent relationship predicates do not include any argument of a non-agentive sort.
The sets of terms of the sorts $R$ and $AgR$ are inductively defined as follows:

i) if $t^{Ag}$ is a term of sort $Ag$, $n \geq 0$, $r$ is a role symbol of sort $(s_j, ..., s_n \rightarrow R)$, and $t_j^{s_j}$ is a term of sort $s_j (j = 1, ..., n)$, then

$$r(t_1^{s_1}, ..., t_n^{s_n})$$

is a role (i.e. a term of sort $R$)

and

$$t^{Ag}:r(t_1^{s_1}, ..., t_n^{s_n}) \quad (t^{Ag}:r, \text{ if } n = 0)$$

is a term of sort $AgR$

(that we say that is indexed by $t^{Ag}$)

ii) if $t^{Ag}$ is a term of sort $Ag$, $n \geq 0$, $r$ is a role symbol of sort $(AgR, s_1, ..., s_n \rightarrow R)$, $u^{AgR}$ is a term of sort $AgR^{32}$, and $t_j^{s_j}$ is a term of sort $s_j (j = 1, ..., n)$, then

$$r(u^{AgR}, t_1^{s_1}, ..., t_n^{s_n})$$

is a role

and

$$t^{Ag}:r(u^{AgR}, t_1^{s_1}, ..., t_n^{s_n})$$

is a term of sort $AgR$ that is indexed by $t^{Ag}$

(That is, a term of the sort $AgR$ always take the form $t^{Ag}:v^R$, for $t^{Ag}$ a term of sort $Ag$ and $v^R$ a term of sort $R$.)

The following abbreviations may lead to shorter and more intuitive descriptions:

i) $t^{Ag}:t^{Ag}$ is an abbreviation of $t^{Ag}:himself$

ii) $r(u^{Ag}, t_1^{s_1}, ..., t_n^{s_n})$ (for $u^{Ag}$ a term of sort $Ag$, and not of sort $AgR$) is an abbreviation of $r(u^{Ag}:u^{Ag}, t_1^{s_1}, ..., t_n^{s_n})$, which (by i)) is an abbreviation of $r(u^{Ag}:himself, t_1^{s_1}, ..., t_n^{s_n})$

The set of formulas (of $L_{DA}$) is inductively defined as follows:

i) if $B$ is a formula of $L$, then $B$ is a formula (a non-modal formula);

ii) the usual Boolean combinations of formulas, are formulas;

iii) if $B$ is a formula, and $x^s$ a variable of a sort $s$ of $L$ (we do not have variables of sort $R$ or of sort $AgR$), then $(\forall x^s)B$ is a formula;

iv) if $B$ is a formula, then $\circ B$ is a formula (a deontic formula);

v) if $t$ is a term of sort $Ag$, or of sort $AgR$, then $E_tB$ is a formula (an action formula with index $t$)

---

32 Normally, it only makes practical sense to consider in this definition that $u^{AgR}$ is a term of sort $AgR$ not indexed by $t^{Ag}$. But we do not need to complicate the language with this requirement.
Semantics

The semantics for the language $L_{DA}$ can be defined through a kind of first-order minimal models $M$, sketched next.

The models take the form $M = <W, I, a set of fe's, fo>$, where:

*) $W$ is a non-empty set of possible worlds.

*) $I$ assigns a standard interpretation structure to the first-order many-sorted language $L$, at each world, satisfying:

i) The domain associated to each sort $s$, a non empty set, is the same at every world (i.e. $I(w, s)=I(w1, s)$, and will be denoted by $D_s$)

ii) $I(w, g): D_{s_1}x...xD_{s_1} \rightarrow D_s$, if $g$ is a function symbol of sort $(s_1,...,s_n)$ (here, and in the next item, some of the sorts $s_1,...,s_n$ may be the sort $Ag$)

iii) $I(w, p) \subseteq D_{s_1}x...xD_{s_1}$, if $p$ is a predicate symbol of sort $(s_1,...,s_n)$

iv) If $c$ is a constant of sort $Ag$, then $I(w, c)=I(w1, c)$ (i.e. these constants are rigid)

v) $I(w, is-himself) = D_{Ag}$

*) We define the sets, $D_R$ and $D_{AgR}$, that are seen as the domains of the sorts $R$ and $AgR$, inductively, as follows:

i) if $r$ is a role symbol of sort $(s_1,...,s_n \rightarrow R)$, $a \in D_{Ag}$, and $a_j \in D_{s_j}$ ($j=1,...,n$), then

\[ r(a_1,...,a_n) \in D_R \quad and \quad a:r(a_1,...,a_n) \in D_{AgR} \]

ii) if $r$ is a role symbol of sort $(AgR,s_1,...,s_n \rightarrow R)$, $a \in D_{Ag}$, $b \in D_{AgR}$, and $a_j \in D_{s_j}$ ($j=1,...,n$), then

\[ r(b,a_1,...,a_n) \in D_R \quad and \quad a:r(b,a_1,...,a_n) \in D_{AgR} \]

*) For each $b$, such that $b \in D_{Ag}$ or $b \in D_{AgR}$, we have a function $fe_b:2^W \rightarrow 2^W$

*) $fo:2^W \rightarrow 2^W$ (and $fo(Z)$, for $Z \subseteq W$, denotes the set of worlds where proposition $Z$ is obligatory).
Given a model $\mathcal{M}$, we define a valuation $\nu$ as a function that applies each world $w$ and each term $t$ of a sort $s$, to an element of $D_s$ (denoted by $\nu(w, t)$) $^{33}$, satisfying:

*) If $t$ is a term of a non-agentive sort, or of sort $Ag$, then $\nu(w, t)$ is defined as it is standard ($\nu(w, c) = I(w, c)$, if $c$ is a constant, and $\nu(w, g(t_1, ..., t_n)) = I(w, g)(\nu(w, t_1), ..., \nu(w, t_n))$, if $g$ is a function)

*) For each variable $x^s$, $\nu(w, x^s) = \nu(w, x^s)$ (i.e. variables are rigid)

*) For the terms of the sorts $R$ and $AgR$, we have (where $r$ is any role symbol and $t, t_1, ..., t_n$ are terms of the appropriated sorts):

\[ \nu(w, r(t_1, ..., t_n)) = r(\nu(w, t_1), ..., \nu(w, t_n)) \]

and

\[ \nu(w, t:r(t_1, ..., t_n)) = \nu(w, t):\nu(w, r(t_1, ..., t_n)) \]

(Note that $\nu(w, t)$ will be independent of $w$, since constants of sort $Ag$, and variables, are rigid; on the other hand, $r$ denotes $r$ in any world)

Truth of formulas is defined with respect to a world $w$ and a valuation $\nu$ in a model $\mathcal{M}$, and can be defined in a more or less standard way for the non-modal formulas. With respect to the modal formulas, we define:

\[ \mathcal{M} \models w, \nu \ E \ B \iff w \in f(e \nu(w, t))(\| B \|_\nu) \]

(for $t$ a term of sort $Ag$, or of sort $AgR$)

\[ \mathcal{M} \models w, \nu \ O \ B \iff w \in f(o(\| B \|_\nu)) \]

where $\| B \|_\nu = \{ w : \mathcal{M} \models w, \nu B \}$

Our models are then constrained in order to validate the desired principles.

**Logical principles (axiomatization)**

The exact logical principles we want are still under research. For each indexed $E$ operator we want the minimal action logic. For $O$ we want at least a classical modal logic of type EC. We

---

$^{33}$ Since the terms of sort $R$ never appear in a formula unless in the scope (as subterms) of a term of sort $AgR$, strictly speaking we would not need to explicitly define $\nu(w, t)$ for the terms $t$ of sort $R$ (it would be sufficiently to define directly the values of the terms of sort $AgR$), neither to define $D_R$. 

probably need some restricted form of the (RM)-rule, since we would like to obtain a classical modal logic of type EC for each indexed $\circ$ (to be introduced next). The axiomatization of the first order component and its interaction with the modal component also needs further research.

With respect to the new action operators proposed, we want at least the following general principle (motivated in section 4):

$$E t : r (t_1, ..., t_n) \models B \rightarrow E t B$$

which becomes valid by imposing the following condition to our models:

$$fe a : r (a_1, ..., a_n) (Z) \subseteq fe a (Z) , \text{ for } Z \subseteq W$$

Moreover, according to our semantics

$$(\forall x^Ag) is\_{-}himself (x^Ag)$$

is valid.

But this is not enough. We would like also to express, as a general principle, that in order to an agent to bring it about something, acting in some quality, he must verify the desired qualifications. For that we define a translation function, $\tau$, that applies each term of sort $AgR^{34}$ to the conjunction of the desired qualifications (the ones that may require authentication). As we will see this function will be also relevant to the attachment of deontic notions to roles.

The function $\tau$ is inductively defined as follows:

i) $\tau (t^Ag : r(t_1^s1, ..., t_n^s)) = is\_r (t^Ag, t_1^s1, ..., t_n^s)$

for $r$ a role symbol of sort $(s_1, ..., s_n \rightarrow R)$

ii) $\tau (t^Ag : r(u^Ag : rl (u_1, ..., u_m), t_1^s1, ..., t_n^s)) =$

$$is\_r (t^Ag, u^Ag, t_1^s1, ..., t_n^s) \land \tau (u^Ag : rl (u_1, ..., u_m))$$

---

$^{34}$ If we decide to admit that the operator $E$ can be indexed by a finite set $X = \{u_1, ..., u_n\}$ of terms of sort $AgR$, then we should extend $\tau$ in the following obvious way: $\tau ((u_1, ..., u_n)) = \tau (u_1) \land ... \land \tau (u_n)$.
for \( r \) a role symbol of sort \((\text{AgR}, s_1, \ldots, s_n \rightarrow \text{R})\) and \( r_1 \) any role symbol

With this function we can then write the desired principle \(^{35}\):

\[
E_t : r(t_1, \ldots, t_n) B \rightarrow \tau(t : r(t_1, \ldots, t_n))
\]

for \( r \) any role symbol and \( t, t_1, \ldots, t_n \) any terms of the appropriated sorts. (We omit here the technical details of how to constrain our models in order to validate this principle.)

Other principles that apply to particular qualities/roles, can also be added to our specifications, if they are relevant for the application at hand. For instance:

\[
E_i : \text{president-of}(k : \text{himself}) B \rightarrow E_k : \text{himself} B
\]

(i.e., using the abbreviations previously introduced:

\[
E_i : \text{president-of}(k) B \rightarrow E_k : k B
\]

**Extending the deontic component**

The language can be extended with non-primitive deontic symbols, following the approach described at section 3, but considering also (similarly) agents acting in a quality (playing a role):

\[ ^* ) \quad P B = \neg O \neg B \quad \text{and} \quad F B = O \neg B \]

\[ ^* ) \quad \text{For} \; t \; \text{a term of sort Ag, or of sort AgR:} \]

\[
O_t B = O E_t B, \quad P_t B = P E_t B \quad \text{and} \quad F_t B = F E_t B
\]

(\text{thus, for instance:} \ O u : r(t_1, \ldots, t_n) B = O E u : r(t_1, \ldots, t_n) B)

\(^{35}\) Another way of approaching this problem would be to consider that the agent relationship predicates, that count as qualities, were of sort \((\text{Ag}, \text{AgR}, s_1, \ldots, s_n)\), with \( n \geq 0 \) and \( s_j \neq \text{Ag} \) and \( s_j \neq \text{AgR} \), for \( j = 1, \ldots, n \), instead of being of sort \((\text{Ag}, \text{Ag}, s_1, \ldots, s_n)\). By imposing the following conditions on our models:

\[ ^* ) \quad (a, b : r_1(b_1, \ldots, b_m), a_1, \ldots, a_n) \in I(w, i s - r) \implies (b, b_1, \ldots, b_m) \in I(w, i s - r) \]

we could then define \( \tau(t : r(t_1, \ldots, t_n)) = i s - r(t, t_1, \ldots, t_n) \), and express the desired principle simply as follows:

\[
E_t : r(t_1, \ldots, t_n) B \rightarrow i s - r(t, t_1, \ldots, t_n)
\]

since, for instance, \( i s - r(t, u : r_1(u_1, \ldots, u_m), t_1, \ldots, t_n) \) would imply (by the previous condition \(^*\)) \( i s - r_1(u, u_1, \ldots, u_m) \).

The two approaches are not completely equivalent, and at this moment we are not sure about which of them better captures our intuitions.
For $X = \{t_1, \ldots, t_n\}$ ($n>1$) a finite set of terms all of sort $Ag$, or all of sort $AgR$:

$O_X B = O_{t_1} B \land \ldots \land O_{t_n} B$ (i.e. informally $\forall x \in X \ O x B$)

$P_X B = P_{t_1} B \land \ldots \land P_{t_n} B$

$F_X B = F_{t_1} B \land \ldots \land F_{t_n} B$

We note that, although we do not exclude the other cases, we will be in practice particularly interested in the deontic formulas where the deontic operators are indexed by terms (or by sets of terms) of sort $AgR$.

**Attaching deontic operators to roles**

And, using the translation function $\tau$, we can also associate deontic operators to roles, in a way that extends naturally the previous definitions:

$O_{r(t_1, \ldots, t_n)} B = (\forall x^{Ag}) (\tau(x^{Ag} : r(t_1, \ldots, t_n)) \rightarrow O x^{Ag} : r(t_1, \ldots, t_n) B)$

(i.e. informally $O \{x^{Ag} : r(t_1, \ldots, t_n) \text{ such that } \tau(x^{Ag} : r(t_1, \ldots, t_n)} B$)

$P_{r(t_1, \ldots, t_n)} B = (\forall x^{Ag}) (\tau(x^{Ag} : r(t_1, \ldots, t_n)) \rightarrow P x^{Ag} : r(t_1, \ldots, t_n) B)$

$F_{r(t_1, \ldots, t_n)} B = (\forall x^{Ag}) (\tau(x^{Ag} : r(t_1, \ldots, t_n)) \rightarrow F x^{Ag} : r(t_1, \ldots, t_n) B)$

for $r$ any role symbol (i.e. a role symbol of sort $(s_1, \ldots, s_{n-1} \rightarrow R)$, with $n \geq 0$, or of sort $(AgR, s_1, \ldots, s_{n-1} \rightarrow R)$, with $n \geq 1$)

With these extensions we can express principles, related with the flow of obligations, and add them to the specifications when convenient. Possible examples are (for $k$ an institutionalized agent):

$P k:\text{himself } B \rightarrow P \text{representative-of}(k:\text{himself}) B$

---

36 Note that we do not introduce sets (neither variables for sets) in our formal language. Thus this universal quantification here should be seen informally, and as an abbreviation of the referred conjunction. The point is that an obligation on a set is seen as an abbreviation of an obligation on every member of the set.

37 In this definition we are using the universal quantifiers of our formal language, and not abbreviations of finite conjunctions. Of course, if, for instance, we define the domain of agents as the set of constants of sort $Ag$, and impose that we have only a finite number of such constants (thus making the domain of agents finite), then we can, at least informally, see such universal quantification as a finite conjunction.

38 We refer again that we do not introduce, in our formal language, sets, nor a language for building and describing sets. The point we want to stress with this "obligation indexed by a set" is that we are informally identifying an obligation attached to a role with an obligation on the set of holders of that role, seen this as an obligation on every member of that set (thus extending in a natural way the previous definition of an obligation on a finite set of terms of sort agent in a role).
\( O_{k:k} B \rightarrow O_{\text{president-of}(k:k)} B \)

(i.e., using the abbreviations previously introduced:

\( P_{k:k} B \rightarrow P_{\text{representative-of}(k:k)} B \)

\( O_{k:k} B \rightarrow O_{\text{president-of}(k:k)} B \)

(Note that the president of an institutionalized agent \( k \) may inherit all the obligations of \( k \), without being the case that the same happens with all the titulars of the various organs of \( k \); that is, the latter formula (schema) may be true, without being true that \( O_{k:k} B \rightarrow O_{\text{titular-of}(k:k)} B \).)

**Section 7. Conclusions and further work**

The problem of collective agency is an important problem when we want to formalize organizations, and agents and organizations interaction. In this paper we have proposed a modal logic approach to this issue. We have argued that once we want (or need) to attribute obligations (permissions or other deontic notions) to a set of agents, then we need to consider a new agent – that we called institutionalized agent – and attribute to him such obligations. But, as we have stressed, once we make that step, we also need to specify how an institutionalized agent interacts with the external world – how the obligations flow from the institutionalized agent to the real agents that support him, and how the actions of the latter count as actions of the former. But, and this is fundamental, an agent may act in many qualities (playing many roles), and it is essential to know in which quality an agent has acted (or intends to act) for three main reasons: to know the effects of the act, its deontic qualification, and authentication issues. For these reasons, we have proposed to extend the action modal operators of the “sees to it” type with an explicit index stating the quality (role) in which the agent has acted, and have studied the main properties desired for this operator. We have also shown how to associate obligations to roles in our formalism, and illustrated how this can be used to express the desired flow of obligations.

This work is far from being finished. The logic proposed needs more research, both with respect to the logic of the deontic operators, and with respect to the interaction between the modal operators and the first-order component. The possible introduction of a notion of “subrole” (within a hierarchy of roles) seems also interesting and needs some more research.

29
A practical language for specifying agents and agent’s interaction is also needed. We are working on these issues at this moment.

Two other interesting research topics are: 1) the study of the “scope operator” referred in Section 5; 2) to model the notion of “juridical consequence”.

Acknowledgments

This work was partially supported by the PRAXIS XXI Program and FCT, as well as by PRAXIS XXI Projects PCEX/P/MAT/46/96 (ACL) and 2/2.1/TIT/1662/95 (SARA). A meeting of José Carmo and Marek Sergot was supported by “acção integrada Luso-Britânica”, nº 8-33/99 (Portuguese-British interchange actions). The authors wish to thank the participants of the last meeting of the project SARA for discussions on some of the topics of this paper, and, in particular, Marek Sergot for further discussions. Finally, the authors also wish to thank the anonymous referees for their comments and criticisms that have helped to improve the effectiveness of this paper.

References


[Chellas 69] B.J. Chellas, The Logical Form of Imperatives, Perry Lane Press, 1969


