Modal Logics for Modeling Agency and Social Interaction (namely organized social interaction)

J. Carmo

(University of Madeira, Portugal)

(Based on joint work with **Olga Pacheco**, Univ. of Minho)

June 23, 2004

Trento

Generic GOAL:

To use modal logics (belief, deontic, action and conditional modal operators) to represent and reasoning about normative concepts, and "structural" concepts (like roles), that are essential to describe (at an abstract level) agents interaction, and that may be seen as the basic building blocks in terms of which organizations are described.

Combination of (modal) logics

We think that a combination of logics of the modal family, like Deontic, Action and Belief logics, together with a conditional ("count as") operator, is a good starting point to analyse, at a convenient abstract level, the fundamental concepts involved in the study of social interaction

In their pioneering work, Kanger, Pörn and Lindahl have already combined deontic and action logics as basic building blocks to describe social interaction and complex normative concepts.

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.

Thus, let us look with some more detail to their logic of action and how to extend it to model the crucial concept of collective agency

Kanger, Pörn and Lindahl action logic

Modal operator (relativized to an agent): E_a

E_a p is read "the agent *a* brings it about that p"
or "agent *a* sees to it that p is the case".

Actions are taken to be relationships between agents and the states of affairs that they bring about, 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.

For instance, using E_a one can express several different positions in which an agent a might be with respect to a certain state of affairs p, such as:

E_a p (did) , $E_a \neg$p (averted) and $\neg E_a$ p $\wedge \neg E_i \neg$p (remained passive)

as well as notions of control of other agents, like:

$E_a E_b$ p (made b do) , $E_a \neg E_b$p (made b avoid), etc.

Moreover, combining E_a 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. by Lindahl.

Criticisms of Castelfranchi (among others): the meaning of E_a p is ambiguous, and iterations like E_a E_b p does not allow to discriminate different causal and intentional structures

For the definition of the logic of the operator E_a, different modal approaches have been taken, providing in general non normal modalities.

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

(T) $E_a\, p \rightarrow p$

(C) $(E_a\, p \wedge E_a\, q) \rightarrow E_a\, (p \wedge q)$

and the RE rule (where $\vdash p \leftrightarrow q$ means $p \leftrightarrow q$ is a theorem of our logic):

(RE) If $\vdash p \leftrightarrow q$ then $\vdash E_a\, p \leftrightarrow E_a\, q$

(besides incorporating the tautologies and the Modus Ponens inference rule)

(The T-schema captures the intuition that if agent *a* brings it about that p, then p is indeed the case; that is, E_a is a "success" operator.)

Thus, we can see these principles as the core of any action logic (of this type), and we will call of *minimal action logic*, a classical modal logic of type ETC (using Chellas classification).

Most of the action logics, of this type, also include the axiom schema:

(No) $\neg E_a\, T$ (where T denotes a tautology)

in order to express that the truth of $E_a\, p$ must imply that the action of agent *a* was necessary to get the state of affairs p

Some extensions and refinements of these action logics have been proposed

Direct action operator:

In 96, Santos & Carmo considered the problem of modelling responsability for a task within an organization, where some members are in positions of management/coordination and others are in operational positions (regarding the production of some task, or part of a task). Although the former are usually made responsible for some task, they only can command or influenced the ones that are able of producing the concrete results.

In order to model such responsability concept, and to deduce its fulfilment and non-fulfilment, they have proposed a distinction between a direct action operator D_a and a non-necessarily direct action operator E_a. Naturally:

$$D_a\, p \rightarrow E_a\, p$$

but it is possible to have

$$E_a\, p \wedge \neg\, D_a\, p$$

(the main distinction between the logics of D_a and E_a regards iterations of the same operator, but we are not enter in details now).

This direct action operator may be useful in other completely different contexts

Attempt operator:

Together with Andrew Jones, they have also proposed a non-necessarily successful action operator H_a, with the following informal meaning:

$H_a\ p$ means that agent *a* attempts to bring about that p

(Naturally, H_a does not verify the T-schema)

What counts as an attempt varies. Besides common sense, there are rules in a legal system, or in an organization (written or from custom) that states that some acts count as an attempt.

In order to express that we can use the "count as" conditional operator introduced by Andrew Jones and Marek Sergot: "In each system (legal system, organization, ...) *s* there exist *meaning rules* that state that some acts, or some state of affairs *count as*, or *are to be classified* as acts, or state of affairs of a different kind (rules that may differ from system to system)"

$(p \Rightarrow_S q)$ means that, for system s (according to the rules operating in such system), p count as q

Using the normal modal operator B_S to denote *s* believes, we have e.g.

$(p \Rightarrow_S q) \rightarrow B_S\ (p \rightarrow q)$ (and so also: $(p \Rightarrow_S q) \rightarrow (B_S\ p \rightarrow B_S\ q)$)

Jones/Sergot use D_S, and not B_S, where $D_S\ p$ is read as follows: “it is incompatible with the constraints operating in system s that p is not the case”. Using $\Rightarrow_S$ we can describe organization's policies regarding what counts as an attempt. A possible policy operating in many organizations s could be:

$$E_a\ O\ E_b\ p\ \Rightarrow_S\ H_a\ E_b\ p$$

(attributions of responsabilities are particular cases of non necessarily successful influences; they will count as an attempt)

as well as (where □ should be seen as meaning ability)

$$E_a\, O\, E_b\, p \;\wedge\; \Box\, E_b\, p \;\Rightarrow_S\; H_a\, p$$

We can also express other possible policies of some organizations, like:

$$H_a\, E_b\, p \;\wedge\; E_b\, p \;\Rightarrow_S\; E_a\, p$$

or even

$$H_a\, p \;\wedge\; p \;\Rightarrow_S\; E_a\, p$$

The attempt operator can be useful in many applications. It may be used to distinguish casual and intentional actions (Attempting involves intention !)

Note that $H_a p$ cannot be reduce to $I_a E_a p$ (an intention or desire to bring about p by myself): I may have an intention to kill someone without exercising that intention (successfully or not).

But we will leave the discussion of this operator, in order to concentrate on the fundamental issue of collective action and agency.

Collective agency

Another natural extension of these logics, considered by Lindahl in 77 consists in allowing the indexing of the operator E by a (finite) set of agents, instead of only by a single agent. Informally:

$E_X\, p$ means that the set of agents described in X jointly see to it the state of affairs p

When we assert E_X p we want to express that the actions of the agents in X cause the state of affairs p; the actions of each of such agents were necessary (or, at least, contributed in a significant way) to the production of p. We may say that the agents described in X jointly cooperate to bring about that p is the case (we leave it here open if such cooperation was intended or not).

Within such extension, we can express some notions of *collective agency*, and define logics where formulas of the form

$$E_{\{a,b\}}\ p \wedge \neg\ E_a\ p \wedge \neg\ E_b\ p \qquad \text{(with } a \neq b\text{)}$$

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.

We can suppose that formulas of the form

$$E_{\{a,b\}}\ p \wedge E_a\ p \qquad \text{(with } a \neq b\text{)}$$

can be also consistent

and there might be even cases where the production of some state of affairs p by an agent *a* "counts as" if the set of agents {a,b} has produced p.

However, by obvious reasons, we reject a general principle of the form

$$E_X\, p \rightarrow E_Z\, p\,, \text{ for } X \subseteq Z$$

HOW TO MODEL ORGANIZATION'S ACTING ?

Organizations are internally “composed” by agents (human agents and possibly also software agents)

But the organizations, themselves, also act and interact in the real world.

Human interaction is normally characterized in terms of normative concepts.

How to apply these concepts to organized collective entities ?

A First Attempt - Combine Deontic and Action Logics

The (Impersonal) Deontic operators usually considered are:

O, meaning obligation

P, meaning permission (P is dual of O: $P = \neg O \neg$)

F, meaning prohibition ($F = O \neg$)

Besides these impersonal deontic operators, we can also conceive similar personal deontic operators, indexing them with an agent:

O_a p : *a* is under an obligation of producing (doing) p

P_a p : *a* is permitted to produce (do) p

F_a p : *a* is forbidden to produce (do) p

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 more natural candidates are:

$O_a p = O E_a p$

$P_a p = P E_a p$

$F_a p = F E_a p$

Some criticisms are made against this option:

First we loose the interdefinability of the relativized deontic operators:

We still get $P_a\, p \leftrightarrow \neg\, F_a\, p$ (and $O_a\, p \rightarrow P_a\, p$, if O satisfies D-schema),

but $F_a\, p \leftrightarrow O_a \neg\, p$ and $P_a\, p \leftrightarrow \neg\, O_a \neg\, p$ are no longer valid

(note that $O_a \neg p = O\, E_a \neg p$ is much stronger than $O \neg E_a p = F_a\, p$).

We do not think that this is a crucial problem.

Second criticism - the "problem of transmission of obligations":

Since (by the T-schema) $\vdash E_a\, E_b\, p \rightarrow E_b\, p$, if we assume a normal modality for O (as in SDL), then the following schema becomes a theorem:

$$O_a\, E_b\, p \rightarrow O_b\, p$$

which is clearly unacceptable.

Moreover, if O satisfies the RM-rule (i.e. that O is closed under implication, in the sense that $\vdash p \rightarrow q$ implies $\vdash O\, p \rightarrow O\, q$), then:

since $\vdash E_a\, p \rightarrow \neg E_b \neg p$ (because of the T-schema)

we get $\vdash O_a\, p \rightarrow F_b \neg p$

It is impossible to express conflicts of obligations between different agents.

Thus, we have two main options. Either to consider that the personal obligations must be primitive, or to define non-normal logics for O where the RM-rule is not verified (even keeping, if we wish, that O satisfies some weaker versions of the RM-rule).

Let us explore the latter option, and 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 e.g. the obligation operator O_X. What is the meaning of O_X p ?

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:

1) $O_X p = \forall_{a \in X} O_a p \quad (= \forall_{a \in X} O E_a p)$

2) $O_X p = \exists_{a \in X} O_a p$

Consider first 2): $O_X\, p = \exists_{a \in X}\ O_a\, p$

Not only it validates $O_X\, p \rightarrow O_Z\, p$, for $X \subseteq Z$, as it is of little 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

$$P_X\, p = \exists_{a \in X}\ P_a\, p \quad \text{and} \quad F_X\, p = \exists_{a \in X}\ F_a\, p$$

then we easily get situations where $P_X\, p$ and $F_X\, p$ are both true

(if $P_a\, p$ and $F_b\, p$, we would get $P_{\{a,b\}}\, p$ and $F_{\{a,b\}}\, p$ are both true).

But, of course, we could define, e.g., F_X through 1) and P_X through 2)

Consider now 1): $O_X\, p = \forall_{a \in X}\, O_a\, p$

Although in many cases we may see an obligation on a set as meaning that all members of the set are under such obligation, 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, as a very simple example, that 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 a1 for player 1, etc.) by: $O_{\{ai:\, i=1,\ldots,11\}}$ five.

Is it correct to infer that every player (or even any player) is under an obligation to mark five goals ? Surely that a situation where

$$E_{\{ai:\, i=1,\ldots,11\}}\, \text{five} \wedge \forall_{i=1,\ldots,11}\, \neg E_{ai}\, \text{five}$$

is true, does not count as a violation of $O_{\{ai:\, i=1,\ldots,11\}}$ five

As another example: If a firm has the obligation to build a house, can we conclude that all the workers of a firm have the obligation to build the (whole) house ?

The previous examples suggest that we follow an approach similar to the one taken above for the definition of O_a p, and define:

3) O_X p = O E_X p

(and similarly for the other deontic operators).

However, this solution is not behind criticisms. Suppose a violation of O_X p occurs (what we might express by O E_X p ∧ ¬ E_X p):

- Who is the responsible of such a violation ?
- What does it mean to simply say that it is the set X that is responsible ?
- 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 essential in the football team example, 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.

Is it meaningful to write O E_X p ?

(There seems to exist situations where P E_X p may be relevant ...)

Of course, we could say that collective deontic operators, like O_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.

First claim: deontic notions, in particular obligations, are only meaningful when associated to agents

If this is true, *the problem relies not on the deontic component, but on the notion of collective agent seen as a set of agents*.

Consider the example of the firm: The firm is different from the set of its workers; the firm has a proper identity (with a name); the set of the workers of the firm can change without changing the identity of the firm !

As soon as we want to assign obligations to a set of agents, we need to do is to consider a new entity – a new agent (organization) that we call of *institutionalized agent* – and associate to him the desired obligations (and

Law seems to support this: some organisations may be subject to obligations, and other deontic qualifications, and be responsible for their violations)

But an organization (an institutionalized agent) cannot act directly ! So how can an organization fulfil its obligations ?

Naturally, someone has to act in the name of the organization ! And this must be specified !

As a brief remark, comparing an organization (an institutionalized agent) with a software agent, we may say that:

- Both can act and be (at least conceptually) subject of obligations
- An organization never acts directly; someone has to act on his name
- But the organization becomes responsible by the effects of the acts made in its name (although an organization cannot go to prison, it can be civil responsible for compensations for damages)
- We may suppose that some software agents can act directly;
- But a software agent cannot be made responsible by the effects of its acts; he always acts in the name of someone (and if something goes wrong it might be important to know whom)

Returning to the organizations, we must specify how an organization interacts with the external world:

- who represents the organization and to what extent

  (who has the power to act on its name)

- and how the obligations attached to an organization become obligations of the "real" agents that supports the organization (i.e. that acts for it):

  otherwise these agents could remain passive (without comitting any violation, but also without fulfilling the organization's duties)

Let us look first to the second issue.

Let us assume that agent *a* made a contract with organization *t*, becoming its President, and for that reason whenever *t* has the oligation to see to it that p, this obligation is transferred to its President.

We could try to express that "flow of obligations" by the formula:

$$O_t\, p \rightarrow O_a\, p$$

But this representation has a problem. Whenever the president of *t* changes we would need to change this "rule"/norm that expresses the flow of the obligation to see to it that p is the case.

Although in an organization there are some tasks / duties / rights that may be assigned to a specific agent (for instance, although it might not be a right of the President of *t* to have a car when on duty, the particular agent *a*, when he made the contract to become President, could have imposed this condition);

normally, an organization has a typified structure formed by some organs, and a set of *stable* rules that associate to these organs the different duties of the organization (that in this way become duties of the members of such organs)

On the other hand, we must be able to express that when the President of *t* sees to it that p, this counts as if the organization has done p (do not forget that it is the organization that is responsible by the non-fulfilment/violation of his obligations).

We could try to express this by a formula of the form:

(*) is-president-of(a,t) $\rightarrow$ (E_a p $\rightarrow$ E_t p)

or, using the "count as" operator (where *s* refers e.g. the legal system)

(*) is-president-of(a,t) $\rightarrow$ (E_a p $\Rightarrow_s$ E_t p)

But this formula fails to represent what is pretended, since the action operator used does not stress the quality (the role) in which *a* has acted when bringing about p.

The sentence *is-president-of(a,t)* expresses that *a* possesses the quality of being President of *t* (agent *a* has that qualification), and so *a* can play the role of President of *t*, but that does not mean that "*a* has acted on that quality (on that role)".

Agent *a* may possess many qualities and produce a similar act acting in different qualities. But the quality in which *a* has acted (or intends to act) is fundamental with respect to two main aspects:

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

The legal effects of an action performed by *a* will depend on the quality *a* has acted (e.g. the legal effects of having a car accident on service, or when driving home, are completely different)

2) *Deontic qualifications:*

An agent may be authorized or even obliged to bring about p acting in some qualities (playing some roles), but not authorized to bring about p in other qualities. (The president of an organization may be permitted to drive the organization car only when he is on duty)

And, so, it may be important to determine the quality in which an agent intends to act: the *authentication* issue.

Thus, we decided to extend the action operators by explicitly indicating in them the quality used by the agent (the role played by the agent) *a* to bring about the state of affairs p, through formulas of the form:

$$E_a \text{ acting as president-of(t) p}$$

or, in a shorter sintax:

$$E_a : \text{president-of(t) p}$$

In order for *a* to act as President of *t, a* must have that qualification:

$$E_a : \text{president-of(t) p} \rightarrow \text{is-president-of(a,t)}$$

Thus, using this operator, we can try to express what we want by writing

$$E_a : \text{president-of(t) p} \Rightarrow_s E_t \text{ p}$$

Special relationships are created between an institutionalized agent (an organization) and other agents (that act for him). To such relationships correspond **roles** that agents can play. When an agent play a role he has different power to act act (not necessarily more - simply different).

The concept of Role

The concept of *role*, and of *acting in a role*, is relevant in many contexts, and not only within the context of the behavior of institutionalized agents. It is a fundamental artifact to understand and describe agent´s acting and interacting.

Roles may be seen as corresponding to qualities that the agents might have, that can be relevant when we describe agents acting and interacting.

We can distinguish two types of qualities:

1) qualities that express properties that an agent might have, independently of the others;

2) qualities that express relationships between two (or more) agents.

In both cases, those properties/relationships may depend on (include) other type of non-agentive 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 established between any two agents).

Examples of 2): worker-of, president-of, administrator-of, friend-of, etc.

Examples of 1): father, policeman, owner-of (owner of a building), etc.

• To act in a role an agent must be qualified to play that role (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 qualification of an agent to play some role must be authenticated somehow: e.g. by the context or through some document.

Different organizations may have different policies regarding the recognition of the role used. E.g. for some organizations, any (relevant) state of affairs p brought about by the president, inside the organization's building, counts as if it was brought about by the president, on that quality:

$$\text{is-president-of}(a,t) \wedge \text{is-in-the-building-of}(a,t) \wedge E_a\, p$$
$$\Rightarrow_t E_{a:\text{president-of}(t)}\, p$$

• Roles are used (mainly within the context of organized collective agency) as a high-level mechanism for structuring the desired behavious, by associating to the roles deontic notions that describe the obligations and permissions for the agents playing such roles. But roles should not be confused with their deontic characterization. Not only two different roles may have similar deontic characterizations, as the deontic characterization of a role may change with time (as a consequence of some actions)

• Associated to a role we have a set of agents (possibly singular): the set of titulars, or holders, of that role (the set of agents that may play that role). But we should not identify the role with such set. First, the set of holders of a role can change. Second, the same set of agents may correspond to the set of holders of distinct roles, 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).

• Contrarily to other authors, we think that we should not see roles as agents. Agents can act, and roles cannot. Of course, when a role has only one holder (like president of an organization), such identification is not dangerous. But what means to say that the role of administrator of *t* has seen to it that p ? Or that the role of father has seen to it that p ?

So, what is a role after all ?

Roles and properties of agents are almost two different faces of a same coin. To each role there corresponds a property that the agents may have (role of father → property of being father). And to each property of an agent we can, *a priori*, associate a role that an agent can play by making use of that property. Naturally, in practice we only associate a role to a property if the fact that someone has that property may be relevant for some of his acts.

• *The role of itself*:

We may assume that when an agent acts, he always act playing some role. Naturally, there are many situations where an agent is acting without exercising any specific relevant property that he has, or any relationship that he has established with other agents. In that cases we can say that the agent is acting in the role of *itself* - a particular role that every agent is qualified to play. We write:

$$E_a : \text{itself } p \qquad \text{or} \qquad E_a : a\ p$$

We can still write $E_a\ p$ to state that agent *a* has brought about p playing some role (that possibly we do not know which was, or we do not care). In such case, we should impose:

$$E_a : \text{role } p \rightarrow E_a\ p$$

We can now be more precise about the effects of acting as President of *t*:

E_a : president-of(t) p $\Rightarrow_s$ E_t : t p (for s e.g. be the legal system)

E_a : president-of(t) p $\Rightarrow_s$ $\neg$ E_a : a p

(although E_a : president-of(t) p $\rightarrow$ E_a p)

Associating deontic notions to roles

Roles are not agents and deontic notions should be applied only to agents. Nevertheless we can (and it is useful to) associate deontic notions to roles that are seen as applying to the holders of such roles.

Omitting technical details, we do that as follows. For each role generator r we associate a predicate symbol is-r, and we define a mapping *qual* that apply each agent *a* and role *r(...)* in the formula *is-r(a,...)*. For instance:

$$\text{qual(a,administrator-of(t))} = \text{is-administrator-of(a,t)}$$

$$\text{qual(a,father)} = \text{is-father(a)}$$

Intuitively, *qual(a, r(...))* denotes the property agent *a* must have in order to be able to play the role *r(...)* (the *qualification* expected to be authenticated when *a* tries to act in that role). Naturally, we impose

$$E_a : r(...)\ p \rightarrow \text{qual}(a, r(...))$$

And then we can define the following abbreviations:

$$O_{role}\ p = (\forall x)(\text{qual}(x,role) \rightarrow O\ E_{x:role}\ p)$$

$$P_{role}\ p = (\forall x)(\text{qual}(x,role) \rightarrow P\ E_{x:role}\ p)$$

$$F_{role}\ p = (\forall x)(\text{qual}(x,role) \rightarrow F\ E_{x:role}\ p)$$

$O_{\text{administrator-of(t)}}\ p$ means that all the administrators of *t* are obliged to see to it that p is the case, *acting in* the quality of administrators of *t*

$P_{\text{administrator-of(t)}}\ p$ means that all the administrators of *t* are permitted of producing p *when acting as* administrators of *t.*

The transmission of the obligation of producing p, from the organization *t* to its President (whatever who is the President) can now be expressed as follows (where $O_{t:t}\ p$ may be seen as $O\ E_{t:t}\ p$)

$$O_{t:t}\ p \rightarrow O_{\text{president-of(t)}}\ p$$

More about the underlying logic (but omitting details)

We assume that an agent may be under contradictory obligations under different roles that he can play, but not in the same role

If we consider O as primitive, then O cannot satisfy the RM-rule. But we need that it satisfies some weaker versions of it. In particular we assume also the following theorems and rules (for r a role):

$O_{a:r}\ p \rightarrow P_{a:r}\ p$ $\qquad$ $O_{a:r}\ p \rightarrow \neg\ P_{a:r}\ \neg p$

$O_{a:r}\ p1\ \wedge\ P_{a:r}\ p2 \rightarrow P_{a:r}\ (p1 \wedge p2)$

If $\vdash p1 \rightarrow p2$ then $\vdash P_{a:r}\ p1 \rightarrow P_{a:r}\ p2$

If $\vdash E_{a1:r1}\ p1 \rightarrow E_{a2:r2}\ p2$ then $\vdash P_{a1:r1}\ p1 \rightarrow P_{a2:r2}\ p2$

Sub-roles

We can have some relations between roles (incompatibility, etc.). A notion of sub-role is particularly useful for specification purposes (to avoid redundancy in the specification of roles).

We will consider that a role r1 is a sub-role of a role r2 if r1 is a more specific than r2

(e.g. President of Department may be seen **(?)** as a sub-role of Member of Department).

In particular a sub-role inherits all the permissions:

$$\text{is-subrole}(r2,r1) \wedge P_{a:r1}\ p \rightarrow P_{a:r2}\ p$$

And, naturally:

$$\text{is-subrole}(r2,r1) \rightarrow (\forall x)\ (\ \text{qual}(x,r2) \rightarrow\ \text{qual}(x,r1)\ )$$

An organization (institutionalized agent) never acts directly !

How to express that ?

We start by noting that a formula like

$E_{a\,:\,a}\ p$ (that is, $E_{a\,:\,itself}\ p$)

does not mean that *a* has produced p acting directly, since it might have been the case that it was some representative of *a* that has produced p, in that quality

But we can attach agents playing roles to the Direct action operator:

$D_{a\,:\,r}\ p$ (agent *a*, acting on the role r, has directly produced p)

$E_{a\,:\,r}\ p$ (agent *a*, acting on the role r, has produced p, directly or not)

and then state that an institutionalized agent never acts directly, by imposing the following principle (for any formula p):

$$(\forall x)\ (\text{is-institutionalized}(x) \rightarrow\ \neg\ D_X\ p)$$

Note, however, that we cannot impose a formula like

$$E_a\ p \rightarrow\ (\exists x)\ D_X\ p$$

(since we have C-schema for E, and it might have been the case that *different parts of* p have been directly produced by distinct representatives of a)

As bridging principles between the action operators, we have:

$$D_{a:r}\ p \rightarrow D_a\ p$$

$$D_a\ p \rightarrow E_a\ p$$

$$D_{a:r}\ p \rightarrow E_{a:r}\ p$$

Using the previous principle, we can describe the direct acts of an agent (when playing some role) using $D_{a:r}$, but continue expressing the effects of such acts on the acts of other agent through the operator $E_{a:r}$

a) **Contracts**

b) Agents are free of establish contracts. Contracts are a source of new roles. In a contract typically roles are attributed to the intervening agents; and the deontic chracterization of those roles is also part of the contract.

Contracts are brought about jointly by two (or more) agents (without being the case that any of them as brought it about alone). Thus, in order to express that some agents have established a contract we propose the use of the action operator indexed by a set of agents playing some roles:

c) $E_{\{a : r1 , b : r2\}}\ C(a,b)$ (where C(a,b) expresses the contract contents)

d) Examples:

e) $E_{\{a:a , b:b\}}\ (O_{a:a}\ pay(25,b) \wedge O_{b:b}\ clean\_house(a) \wedge ...)$

f) $E_{\{a:president\text{-}of(t) , b:b\}}$ (is-administrator-of(b,t) $\wedge$ $O_{t : t}$ pay(100,b) $\wedge$ ...)

Contracts also include in general conditional obligations (and/or conditional permissions), namelly describing the effects of the fulfillment or violation (unfulfillment) of other obligations in the contract

g)

h) The obligations, or other normative concepts, associated in the contracts to the roles are attributed to the concrete agents in the respective roles

i)

j) On the contrary, we associate directly obligations, permissions, etc. to the roles that constitute the structure of an institutionalized agent, and only a *posteriori* they apply to the holders of such roles; thus, these can change, but the deontic characterization of the *structural roles* remains

k) Conclusions and further work

The problem of collective agency is an important problem when we want to formalize organizations, and agents and organizations interaction.

We have argued that once we want (or need) to attribute obligations (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 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.

We have also stressed that 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. 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.

Much remains to be done !