

RESPONSIBILITY FOR ACTION IN ORGANISATIONS: A FORMAL MODEL

Filipe A. A. Santos	Andrew J. I. Jones	José M. C. L. M. Carmo
<i>Dept. of Information Systems</i>	<i>Dept. of Philosophy & NRCCL</i>	<i>Dept. of Mathematics</i>
<i>ISCTE</i>	<i>University of Oslo</i>	<i>IST</i>
<i>Av. das Forças Armadas</i>	<i>PB 1024 Blindern</i>	<i>Av. Rovisco Pais</i>
<i>1600 Lisboa Codex</i>	<i>N-0315 Oslo</i>	<i>1096 Lisboa Codex</i>
<i>Portugal</i>	<i>Norway</i>	<i>Portugal</i>
<i>fas@iscte.pt</i>	<i>andrew.jones@filosofi.uio.no</i>	<i>Jose.Carmo@math.ist.utl.pt</i>

ABSTRACT

Modal action operators are combined with a deontic logic and a conditional logic, and applied to the analysis of some simple organisational structures. The characterisation of the notion of responsibility, and of the conditions under which an organisation recognises that an agent has fulfilled the responsibilities assigned to him, are the main focus of interest. It is argued that, in this context, distinctions should be articulated between *direct* and *indirect* action, and between *successful* and *not necessarily successful* action.

The general line of approach taken here to the modal logic of action follows the tradition established by Kanger and Pörn.

1. THE APPROACH

The principal aim of this paper is to show that a set of action and deontic modalities, supplemented by a conditional operator, can provide the basic building blocks to be used in the characterisation of some central organisational notions; the particular emphasis here will be on the concept of *responsibility for action*. By *organisation* we mean a society of agents,

comprising human or artificial agents (or both), whose rule-governed interactions are aimed at some specified tasks or goals. We adopt the perspective proposed in (Normatics, 1991), viewing organisations as instances of *normative systems*, and describing agents' interactions in terms of what those agents are permitted or obliged to do (allowing also the possibility that their behaviour may deviate from the ideal), and in terms of other complex normative relations between them. One principal advantage of adopting this perspective is that it enables us to capitalise on a lot of the work that has been done on the application of formal-logical techniques to the analysis of law and legal systems. If we wish to model real societies of agents, or if we wish to design systems which automate some aspects of their activities, it is essential that such concepts as *right*, *permission*, *obligation*, *authority*, *authorisation*, *responsibility* and *delegation*, are precisely understood and defined, and not simply treated informally (see, e.g., Jones et al., 1992, 1996).

We are, of course, aware of the kinds of theoretical models that have been proposed in *Organisation Theory* (see, e.g., Tosy, 1994; Robbins, 1987; Hodge et al., 1988). In a future work (Santos, in preparation) a systematic overview of some main theoretical trends in existing Organisation Theory will be presented, and they will be compared to the perspective here advocated. For the moment, we simply want to register the opinion that too little has so far been done in Organisation Theory to supply precise qualitative models of norm-governed interaction, and we shall seek to show that some techniques drawn from modal logic may help to fill that gap.

Our approach has been the following: we take advantage of previous contributions of applied modal logic to the representation of norm-governed action (see, e.g., Kanger, 1971, 1972; Pörn, 1970, 1977, 1989; Lindahl, 1977); we confront their expressive power with further concepts relevant to the specification of organisations; and, where necessary, we propose additional modalities in order to cope with those concepts. (Some of the ideas discussed here were previously presented in Santos et al., 1996a, 1997.)

An overview of the rest of the paper follows. In Section 2, we make a brief description of the logical context adopted as our starting point for representing social interaction. In Section 3, we describe the kinds of characterisation problems concerning organisations/ organised interaction which we want to be able to solve. By appeal to these problems and by confronting the expressive power of the logics of action proposed by other authors, we discuss the need for the introduction of other action modalities to articulate, first, the distinction between a “direct” and an “indirect” agency concept, and, second, the distinction between “successful” and “not necessarily successful” action. In Section 4, we describe the axiomatic and model-theoretic specifications of these modalities. In Section 5, we apply the action operators to some concrete issues arising in the description of organisational activities. The problem of “transmission of agency”, and the analysis of the conditions under which an organisation recognises that an agent has fulfilled his responsibilities, are focal points. Conclusions and directions for future work appear in Section 6.

2. THE BACKGROUND

In their pioneering work, Kanger, Pörn and Lindahl have combined deontic logic and action logic as basic building blocks to describe social interaction and complex normative concepts (see, e.g., Kanger, 1971, 1972; Pörn, 1970, 1977, 1989; Lindahl, 1977). 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 relativised modal operator, here designated by E_i , where expressions of the form E_iA are read “the agent i brings it about that A ” or “agent i sees to it that A is the case”. The formal-logical development of this approach is due to Pörn (Pörn, 1970), with more refined versions presented in (Kanger, 1972; Pörn, 1977; Elgesem, 1993). An important feature of these logics is that actions are taken to be relationships between agents and the states of affairs that they bring about, setting aside temporal aspects. (For a brief overview of action logics in this tradition, see Santos et al., 1996a.)

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

$$(T) \quad E_iA \rightarrow A$$

$$(C) \quad (E_iA \wedge E_iB) \rightarrow E_i(A \wedge B)$$

and the rule of inference:

$$(RE) \quad \text{If } \vdash A \leftrightarrow B \text{ then } \vdash E_iA \leftrightarrow E_iB$$

(The (T) schema captures the intuition that if agent i brings it about that A , then A is indeed the case; that is, E_i is a “success” operator.)

This approach to the logic of action offers an expressive power rather different from that of action logics based on dynamic logic (see e.g. Harel, 1979, 1984); for instance, it facilitates the expression of the different atomic positions an agent might be in with respect to a particular state of affairs A : E_iA , $E_i\neg A$ and $\neg E_iA \wedge \neg E_i\neg A$ (the last of these meaning “ i remains passive with respect to A ”); and it readily affords a means of characterising cases of interpersonal control, for instance by means of expressions of the form E_iE_jA (*cf.* Pörn, 1970, 1977). (For a further brief comparison between these two types of approach to the logic of action, see, for instance, Brown et al., 1996.)

The importance of this operator to the characterisation of norms is brought out by the distinctions it affords when combined with a deontic operator, as has been shown in, for instance, the work of Kanger (Kanger et al., 1966) and

Lindahl (Lindahl, 1977). As an example of the kind of analysis this combination of operators provides (assuming the adoption of Standard Deontic Logic - SDL), we list the seven *normative one-agent act-positions* (where the operator O represents “it is obligatory that” and its dual P represents “it is permitted that”). For any agent i , and for any state of affairs A , precisely one of the following positions obtains (*cf.* Jones et al., 1992):

- (T1) $PE_iA \wedge PE_i\neg A \wedge P(\neg E_iA \wedge \neg E_i\neg A)$
- (T2) $PE_iA \wedge O\neg E_i\neg A \wedge P(\neg E_iA \wedge \neg E_i\neg A)$
- (T3) $PE_iA \wedge PE_i\neg A \wedge O(E_iA \vee E_i\neg A)$
- (T4) $O\neg E_iA \wedge PE_i\neg A \wedge P(\neg E_iA \wedge \neg E_i\neg A)$
- (T5) OE_iA
- (T6) $O(\neg E_iA \wedge \neg E_i\neg A)$
- (T7) $OE_i\neg A$

The *Theory of Normative Positions* has been further developed to take into account, for instance, relationships between two agents, giving thirty-five *normative individualistic two-agent act-positions*, and to cover also situations where the obligations apply collectively, but not necessarily individually, giving one hundred and twenty-seven *collectivistic two-agent act-positions*.

The usefulness of these distinctions is well illustrated by their contribution to the analysis of Hohfeld’s “fundamental legal conceptions” (Hohfeld, 1923), to be found in the work of Kanger, Pörn and Lindahl (the latter also dealing with questions related to change of normative relations). But, in our view, the interest of these theories is certainly not limited to their contributions to legal philosophy. In any norm-governed system, legal or otherwise, if it is felt essential to be very precise about the particular normative positions the agents concerned occupy, then formal theories of these kinds provide useful tools. For instance, one important application of the Kanger-Lindahl theory concerns the specification of computer security requirements, as has been shown in (Jones et al., 1992; Morris et al., 1991).

Pörn was one of the very first to appreciate the potential contribution of this approach to the characterisation of organised interaction (see Pörn, 1970, 1971, 1977, 1989). Pörn combined the previously mentioned action and deontic concepts with other modalities, such as a doxastic operator B , yielding interesting characterisations of *intentional action*, *control*, *influence*, *communication*, among other notions, and using them, in combination with other mathematical models (drawn, for example, from games theory and automata theory), to discuss the structure of social systems. It is a great pity, in our opinion, that Pörn’s work has not had a greater impact on mainstream Organisation Theory.

3. RESPONSIBILITY FOR ACTION IN ORGANISATIONS

All organisations, *formal* or *informal*, contain an underlying structure that stipulates the allocation of tasks to individuals, their interaction patterns and their coordination mechanisms. This structure is intended to provide the work distribution necessary to attain organisational goals. Work divided in this way is grouped into different operational or management/coordination positions, that are allocated to individuals through the assignment of responsibilities. This organisational concept of responsibility usually involves obligations that members of the organisation must fulfil, in accordance with procedures, policies and strategies. On the one hand, agents in operational positions fulfil their obligations by acting according to their capabilities, and on the other hand, agents in management positions fulfil their obligations by ensuring that certain results are obtained via other agents, by using existing organisational procedures or by creating new ones, by delegating responsibilities, etc. Although the situation may vary according to the degree of formalization adopted in a given organisation, a common feature of these obligations is that they often refer to the specific results that must be achieved within the organisation, *without entering into details concerning the particular actions that the agents themselves must perform in order to achieve these results*. In fact, the concrete actions to be performed to fulfil those obligations depend on the underlying structure of the organisation (e.g., power structure, task decomposition), and on the specific circumstances (e.g. availability of specific agents and resources).

It is commonly the case in organisations that an agent (a manager, say) is made responsible for securing the realisation of some state of affairs which he is either not capable of bringing about himself (he perhaps lacks the required skills), or is not permitted to bring about himself (he is perhaps not a member of the relevant Trade Union, or maybe the organisation does not consider it to be *his* job to perform the tasks concerned). Clearly, the manager's responsibility involves a requirement that he exercise power and influence over others in an effective way, getting them to perform the tasks necessary for reaching the goal that he (the manager) is obliged to secure. This indicates that, in order to characterise this notion of responsibility, it will be essential to define a notion of "indirect agency", to capture cases of agency in which an agent secures the production of some state of affairs A without necessarily bringing it about that A himself.

In what follows we propose to use the word *ensure* to represent the "indirect" action concept. On the other hand, we retain the expression "bring it about" for an action concept (made explicit in Section 4) which may be called "direct" in at least the following sense: an agent *i* brings it about that A only if he does not bring it about that some *other* agent *j* brings it about that A.

We use a monadic operator G_i to represent the "indirect" agency concept, keeping E_i to stand for our notion of "direct" agency; we read expressions of the form G_iA as "agent *i* ensures that A". Using this operator in combination with a suitable

deontic obligation operator O , we now let expressions of the kind “agent i is responsible for A ” be represented by OG_iA (i.e., responsibility for A amounts, on this view, to an obligation to ensure that A).

Like the direct action modality, the indirect action modality is considered to be “successful”. Thus we may consider that such expressions as $OG_iA \wedge G_iA$ and $OG_iA \wedge \neg G_iA$ represent an acceptable characterisation of “fulfilment” and “non-fulfilment” of responsibilities, respectively. However, “non-fulfilment” represented in this way does not bring out the degree of participation of the responsible agent in the organisational activity required to obtain A , and his participation, we argue, can be an important element to be taken into account when analysing different non-fulfilment situations. Consider, for instance, a situation in which a responsible agent has ordered his subordinate to perform a given task, but where the subordinate does not follow the order. This situation is obviously different from the situation in which the responsible agent does nothing at all. These two situations are likely to have rather different consequences within a given organisation ! The introduction of a distinction between “successful” and “not necessarily successful” action will allow us to cope with these situations. By a “not necessarily successful” action we simply mean an action performed by an agent in an attempt to achieve some goal, where that attempt may fail to be successful.

We propose to use a monadic operator H_i for representing “not necessarily successful” actions, where expressions of the form H_iA are read “agent i attempts to make it the case that A ”. The two types of non-fulfilment of responsibility mentioned in the previous paragraph may now be represented as follows: $OG_iA \wedge \neg H_iA$ and $OG_iA \wedge \neg G_iA \wedge H_iA$.

In addition, we can now also express weaker types of obligations/responsibilities of the form OH_iA , although they are likely to arise within organisations only in circumstances in which there are strong reasons to expect that it will be very difficult to obtain A ¹. On the other hand, obligations of the form $O\neg H_iA$ may occur more often, since they represent the not unusual situation in which even an attempt is forbidden (e.g., agent i is forbidden to attempt to access his medical record). Thus, $O\neg G_iA \wedge O\neg H_iA$ and $O\neg G_iA \wedge \neg O\neg H_iA$ represent two relevantly different situations. (As would be expected, the logical properties of the modalities O , G_i and H_i will make redundant the first conjunct in $O\neg G_iA \wedge O\neg H_iA$.)

By combining/iterating our action modalities, a number of interesting distinctions may be drawn, for instance in regard to concepts of interpersonal control with respect to A , e.g., E_iE_jA and E_iG_jA for “direct” control and G_iE_jA and G_iG_jA for “indirect” control. These are cases where we may say that agent i exerts a “successful influence” on agent j , whose action, in turn, is also successful. Furthermore, E_iH_jA and G_iH_jA also represent species of “successful influence” on agent j (in the sense that, in both cases, what i does secures the result that j attempts to make it the case that A). By contrast,

¹ Not surprisingly, a common feature of organisational responsibilities is that they are described in a way that indicates that responsible agents are required to succeed !

$H_i E_j A$, $H_i G_j A$ represent “not necessarily successful influences” of agent i on j . These are just a few illustrations of the expressive power afforded by these modalities.

In the next Section we make more precise the intended interpretation of the action notions, in part by exploring in more detail their interrelationships.

4. THE NEW ACTION OPERATORS: FORMAL AND SEMANTICAL FEATURES

Starting from the logical properties of the “direct” action operator E_i , we propose the use of a propositional classical system of type ECT (according to the classification of Chellas, 1980) including also the schema (No), i.e., a system based on classical propositional logic, and with the following axiom schemas:

- (T) $E_i A \rightarrow A$
- (C) $(E_i A \wedge E_i B) \rightarrow E_i(A \wedge B)$
- (No) $\neg E_i \text{True}$

and the rule of inference:

- (RE) If $\vdash A \leftrightarrow B$ then $\vdash E_i A \leftrightarrow E_i B$

The additional schema (No) captures the idea that the truth-conditions of $E_i A$ involve a negative component (*cf.* Pörn, 1977; Belnap, 1991), requiring that if i had not acted as he did, A might have failed to obtain. Clearly, where A is a tautology, this condition is not met. (But see Elgesem, 1993 for a critical discussion of the proper interpretation of this negative component.)

The same intuitions apply with respect to the formal properties of the operator G_j , i.e., we also adopt the axiom schemas:

- (T) $G_i A \rightarrow A$
- (C) $(G_i A \wedge G_i B) \rightarrow G_i(A \wedge B)$
- (No) $\neg G_i \text{True}$

and the rule of inference:

(RE) If $\vdash A \leftrightarrow B$ then $\vdash G_i A \leftrightarrow G_i B$

Thus, we may say that these schemas form the kernel of our “successful” agency concepts (*cf.* Santos et al., 1996a).

However, the main difference between the operators E and G concerns interaction between different agents. The idea of “indirect control” discussed in the previous section suggests the additional schema

(GGG) $G_i G_j A \rightarrow G_i A$

This schema has the following intuitive reading: “whenever agent i ensures that agent j ensures that A, agent i also ensures that A” and marks the key difference between a “direct” and an “indirect” action concept. Since $G_i G_j A \rightarrow (G_i A \wedge G_j A)$ (by (T) and (GGG)), each agent involved in $G_i G_j p$ ensures that p.

On the other hand, we adopt the schema

(EE¬E) $E_i E_j A \rightarrow \neg E_i A$ (for $i \neq j$)

which makes explicit the idea that $E_i A$ implies that agent i brings it about that A *directly*, at least in the sense that his bringing it about that A is incompatible with his bringing it about that some other agent j brings it about that A.² However, we do not adopt the schema $E_i G_j A \rightarrow \neg E_i A$ ($i \neq j$), since it conflicts with the following situation where there is a kind of “circular” chain of control: agent i sees to it that j ensures A, and j ensures A by seeing to it that i sees to it that A. This situation, we think, can consistently be described by $E_i G_j A \wedge E_j E_i A$.

Note that we may consistently maintain $E_i E_j A \rightarrow E_i A$. However, this schema (which is of course an instance of the (T) schema) provides grounds for rejecting schema (4) for the operator E, i.e., the schema $E_i A \rightarrow E_i E_j A$. The reason is that schema (4), together with (T), (RE) and (EE¬E), entails the following: $E_i E_j A \rightarrow \neg E_i E_j A$.³

² A stronger notion of *direct* agency - which in virtue of (EG) below entails (EE¬E) - would be captured by adoption of:

(GE¬E) $G_i E_j A \rightarrow \neg E_i A$ (for $i \neq j$)

The consequences of adopting this or other alternative candidates for the characterisation of *direct* agency are a topic for further enquiry. Adoption of (EE¬E) is sufficient for our present purposes in marking a contrast between the G and E operators. But a great deal more remains to be said about how direct and indirect agency concepts might be distinguished.

³ Suppose $E_i E_j A$. If both (4) and (T) are adopted for E, then $E_j A \leftrightarrow E_j E_j A$. Then, by (RE), we have $E_i E_j A \leftrightarrow E_i E_j E_j A$. But, by (EE¬E), $E_i E_j E_j A \rightarrow \neg E_i E_j A$.

It should be noted that the logical characterisations of the action notion provided by the authors mentioned in section 2 do not bring out the points we are here emphasising. The logics in (Kanger, 1972; Pörn, 1977; Elgesem, 1993) have counterparts to neither (GGG) nor (EE¬E) and thus we may say that they remain neutral with respect to these important aspects of agency. Just the counterpart to (GGG) is adopted in (Pörn, 1970). However, Pörn there employs a normal system (i.e., a system closed under the rules (RM) and (RN)) and thus the logic contains some apparently counter-intuitive results. Accepting a normal system for E, for instance, would allow us to derive the schema $E_i\text{True}$, and the operator would be closed under logical consequence, so that - for example - if i sees to it that j knows that p, it follows that i sees to it that p. Problems relating to consequential closure are the principal reason for our rejection of the schema (M), for acceptance of (M), together with (C) and (RE), generates (RM).

Although we have opted for a non-normal interpretation of the indirect action modality G, the reasons for *that* choice are perhaps less compelling than in the case of the E operator. Were (RM) accepted for E then, since the schema (T) is adopted, E_iA would be entailed by E_iE_jA , in conflict with the main idea underlying the characterisation of the “direct” action notion, i.e., the schema (EE¬E).

With respect to the relationships between E and G, the following schema reflects the idea that “bringing it about is a particular case of ensuring”:

$$(EG) \quad E_iA \rightarrow G_iA$$

With respect to the operator H_i , we propose the use of a propositional classical system of type EC (according to the classification of Chellas, 1980), i.e., a system containing the following axiom schema:

$$(C) \quad (H_iA \wedge H_iB) \rightarrow H_i(A \wedge B)$$

and the rule:

$$(RE) \quad \text{If } \vdash A \leftrightarrow B \text{ then } \vdash H_iA \leftrightarrow H_iB$$

Comparing this proposal with the previous ones, the absence of the (T) schema is of course justified by the intuitions discussed with respect to a “not necessarily successful” action concept. We do not adopt (No) for H, since perhaps some sense can be made of the idea that an agent might attempt to make true that which is logically true...but very little turns on that choice as regards the role to be played by the H modality.

The logic of H will parallel the logic of G - and differ from that for E - in as much as the following schema is adopted:

$$(HHH) \quad H_i H_j A \rightarrow H_i A$$

with the reading : “whenever agent i attempts to make it the case that agent j attempts to make it the case that A, agent i also attempts to make it the case that A”.

With respect to the relationship between H and the successful-action operators, we propose the following schema, reflecting the idea that “ensuring is a successful attempt to make it the case”:

$$(GH) \quad G_i A \rightarrow H_i A$$

The logical principles of the H operator are quite simple. However, further principles may be discussed in relation to some possible additional assumptions. For instance, imposing “rationality” on the behaviour of each agent would suggest the adoption of the (D) schema, i.e., $H_i A \rightarrow \neg H_i \neg A$. Moreover, if it were to be assumed that the efforts of each agent were coordinated in a rational fashion, then it would be natural to adopt the schema: $H_i A \rightarrow \neg H_j \neg A$. However, it is often - perhaps usually - the case that rationality assumptions of these kinds cannot be made with respect to actual, as opposed to ideal, agents.

As regards further links between the three action concepts, we propose to adopt the following bridging principles:

$$(EEEG) \quad E_i E_j A \rightarrow E_i G_j A$$

$$(EGEH) \quad E_i G_j A \rightarrow E_i H_j A$$

$$(GEGG) \quad G_i E_j A \rightarrow G_i G_j A$$

$$(GGGH) \quad G_i G_j A \rightarrow G_i H_j A$$

$$(HEHG) \quad H_i E_j A \rightarrow H_i G_j A$$

$$(HGHH) \quad H_i G_j A \rightarrow H_i H_j A$$

Although we rejected, for each of the three modalities, the rule (RM) of closure under logical consequence, each of the above six principles exhibits a restricted form of closure. For instance, the schema (EEEG), which states that “whenever i brings it about that j brings it about A, i also brings it about that j ensures A”, expresses the closure of the E operator with respect to the principle (EG); and, to mention just one other case, (HGHH) expresses the closure of the H operator with respect to the principle (GH).

From these schemas, and those adopted previously, we may derive, among others, the schemas $G_i E_j A \rightarrow G_i A$, $H_i E_j A \rightarrow H_i A$, $H_i G_j A \rightarrow H_i A$.

We have so far characterised the three modalities and their interrelationships in terms of the axiom schemas and rules of inference that we wish to adopt. We next give, in barest outline, a semantical characterisation.

For the semantic characterisation of the operators E, G and H, we adopt the model $M = \langle W, f_1, \dots, f_n, g_1, \dots, g_n, h_1, \dots, h_n, P \rangle$, where: W is a set of possible worlds; P is a function assigning each propositional atom p_i to the set of elements of W where p_i is true; and each f_i , g_i and h_i ($i = 1 \dots n$) is a function with signature $2^W \rightarrow 2^W$, where $f_i(X)$ (resp., $g_i(X)$ and $h_i(X)$) is the set of worlds where agent i *brings about* (the state of affairs described by) X (resp., the set of worlds where agent i *ensures* X and the set of worlds where agent i *attempts to make it the case that* X).

Truth at a world in a model is defined in the standard way for non-modal expressions, and for expressions of the form $E_i A$, $G_i A$ and $H_i A$ by the following rules (where $\|A\|$ is the truth-set of the sentence A):

- (def.e) $M, \alpha \models E_i A$ iff $\alpha \in f_i(\|A\|)$
 (def.g) $M, \alpha \models G_i A$ iff $\alpha \in g_i(\|A\|)$
 (def.h) $M, \alpha \models H_i A$ iff $\alpha \in h_i(\|A\|)$

To validate the schemas discussed above, the following constraints on the model M are imposed: for all X and Y (subsets of W) and for all agents i, j :

- (te) $f_i(X) \subseteq X$
 (ce) $f_i(X) \cap f_i(Y) \subseteq f_i(X \cap Y)$
 (noe) $f_i(W) = \emptyset$
 (tg) $g_i(X) \subseteq X$
 (cg) $g_i(X) \cap g_i(Y) \subseteq g_i(X \cap Y)$
 (nog) $g_i(W) = \emptyset$
 (ggg) $g_i(g_j(X)) \subseteq g_i(X)$
 (ee¬e) $f_i(f_j(X)) \subseteq W - f_i(X)$, for $i \neq j$
 (eg) $f_i(X) \subseteq g_i(X)$
 (ch) $h_i(X) \cap h_i(Y) \subseteq h_i(X \cap Y)$
 (hhh) $h_i(h_j(X)) \subseteq h_i(X)$
 (gh) $g_i(X) \subseteq h_i(X)$

- (eeeg) $f_i(f_j(X)) \subseteq f_i(g_j(X))$
- (egeh) $f_i(g_j(X)) \subseteq f_i(h_j(X))$
- (gegg) $g_i(f_j(X)) \subseteq g_i(g_j(X))$
- (gggh) $g_i(g_j(X)) \subseteq g_i(h_j(X))$
- (hehg) $h_i(f_j(X)) \subseteq h_i(g_j(X))$
- (hggh) $h_i(g_j(X)) \subseteq h_i(h_j(X))$

We offer here no attempt to give a deeper intuitive account of the informal interpretation of these semantical conditions, which must remain as a task for a more detailed exposition of this language. We have, however, tried to indicate its expressive power, and further illustrations of this will follow in later sections.

For those interested in the generation of act-positions, we should mention that by incorporating both the operators E_i and G_i , we get five one-agent act-positions, instead of the three described in (Lindahl, 1977; Jones et al., 1993) for the single operator E_i :

- (A1) $E_i A$
- (A2) $E_i \neg A$
- (A3) $G_i A \wedge \neg E_i A$
- (A4) $G_i \neg A \wedge \neg E_i \neg A$
- (A5) $\neg G_i A \wedge \neg G_i \neg A$

Furthermore, taking all three operators E_i , G_i and H_i (and adopting the schema (D) for H_i), we get seven one-agent act-positions:

- (A1) $E_i A$
- (A2) $E_i \neg A$
- (A3) $G_i A \wedge \neg E_i A$
- (A4) $G_i \neg A \wedge \neg E_i \neg A$
- (A5) $H_i A \wedge \neg G_i A$
- (A6) $H_i \neg A \wedge \neg G_i \neg A$
- (A7) $\neg H_i A \wedge \neg H_i \neg A$

Without the “rationality” assumption provided by schema (D) for the operator H, we would get thirteen positions.

5. THE EXERCISE OF INFLUENCE AND THE FULFILMENT OF RESPONSIBILITIES

In order to illustrate the inference capabilities of the logic proposed above, let us consider the following example of an organisation with three agents a , b and c , where the agents have just the capabilities indicated in Figure 1. We further assume that agent a has effective influence over agents b and c , with respect to $p1$ and $p2$, respectively.

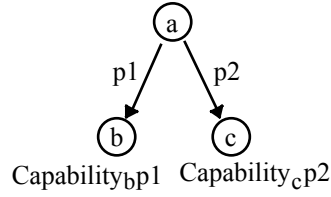


Figure 1.

Within this simple structure, there is just one possibility for achieving $p1 \wedge p2$: each agent must act according to his capabilities. Thus, in order that a can ensure $p1 \wedge p2$, he must exercise successful influence on agents b and c to ensure $p1$ and $p2$, respectively. In fact, the logical system provides the following deduction:

$$\{E_a G_b p1, E_a G_c p2\} \vdash G_a(p1 \wedge p2)$$

Now let us suppose that agent a exercises influence on agents b and c with a view to ensuring $p1$ and $p2$, but that agent c does not succeed in obtaining $p2$, whereas agent b succeeds in achieving $p1$. We may say in this case that agent a has just ensured $p1$, in his attempt to secure both $p1$ and $p2$; the following deduction is forthcoming in the logic:

$$\{E_a G_b p1, E_a H_c p2\} \vdash G_a p1 \wedge H_a(p1 \wedge p2)$$

Finally, consider the following variation of this scenario: suppose that agent c did not perform any action. This case will generate the same conclusions as the previous one, although there is now of course a difference with respect to the influence exercised by agent a on c : agent a attempted to get c to ensure that $p2$, but was not successful. There is just one possible candidate for representing a 's attempted influence in terms of the combinations of the action operators introduced above: $H_a G_c p2$. The logical system yields the deduction:

$$\{E_aG_{bp1}, H_aG_{cp2}, \neg H_{cp2}\} \vdash G_a p1 \wedge H_a(p1 \wedge p2)$$

The hypotheses used in these deductions represent the situations concerned with a certain degree of abstraction. An interesting question that naturally arises is how these abstract notions might have been derived.

If we were to assume that within this organisation *a* exercised his influence by means of “attributions of responsibilities”, that would lead us to propose the following set of hypotheses as a slightly more concrete description of the above situations: $\{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, E_{cp2}\}$, $\{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, H_{cp2}, \neg E_{cp2}\}$ and $\{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, \neg H_{cp2}\}$, respectively. However, even if we were to accept:

$$(\text{attrib.resp}) \quad E_iOG_jA \rightarrow H_iG_jA$$

which says that “attributions of responsibility” are particular cases of “not necessarily successful influences”, that would not enable us to derive the previous conclusions from these sets of hypotheses.

The reason for this, we suggest, is that we have no principle of “transmission of agency” relating the actions of influence with the other actions within the “chain of successful influences” underlying an “indirect” action. Were we to accept, for instance, the principle

$$(\text{trans.resp}) \quad (G_iOG_jA \wedge G_jA) \rightarrow G_iA$$

to represent the “transmission of agency” involved in most institutionalised organisations (where influences are exercised by means of assignments of responsibilities), then that would allow us to derive the above conclusions from the more concrete descriptions we offered of the previous situations, i.e.:

$$\begin{aligned} \{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, E_{cp2}\} &\vdash G_a(p1 \wedge p2) \\ \{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, H_{cp2}, \neg E_{cp2}\} &\vdash G_a p1 \wedge H_a(p1 \wedge p2) \\ \{E_aOG_{bp1}, E_aOG_{bp2}, E_{bp1}, \neg H_{cp2}\} &\vdash G_a p1 \wedge H_a(p1 \wedge p2) \end{aligned}$$

Note that, within organisations, (trans.resp) is usually accepted in the sense that “whenever agent *i* has influenced agent *j* (using his official power of influence) to ensure *A*, and agent *j* has ensured *A*, then for the organisation this *counts as*

agent i 's ensuring A ". These modest examples illustrate, we believe, the power of the logical representation with respect to uncovering some important implicit assumptions regarding the exercise of influence.

However, the principle (trans.resp) should not be seen as a general "transmission of agency" principle, for the following reason: it can happen that agent i has exercised an influence on agent j to ensure A , and that agent j has ensured A *but not because of i 's influence*, but for some other reason. Is it not then doubtful whether one could conclude that i has ensured that A ?. Our opinion is that the answer to this question will depend on the "transmission of agency" principles *adopted by the organisation*. For instance, some organisations adopt as a part of their usual practice that responsibilities are deemed to have been met provided simply that the intended results have been obtained. We may adopt :

$$\text{(trans.who.cares)} \quad (OG_jA \wedge A) \rightarrow G_jA$$

as a correct representation of this organisational position.

Since these "transmission of agency" principles may vary from organisation to organisation, depending on the policies they adopt, they obviously cannot be considered to be general logical principles.

The policies adopted by an organisation reflect in part how it interprets the significance of the actions performed by its agents - how it interprets the meaning of these actions, or what they "count as". The idea of considering a "meaning system" as an important element of analysis for organisations is not new, and has been adopted by recent "institutional" theorists (see, e.g., Scott et al., 1994). In these theories, agents (either individual or collectives) and their actions (either individual or collective) are seen as social constructs whose meaning depends on symbolic elements of various kinds.

So, we may further develop our action notions by considering the interpretation that an organisation gives to some of its actions. For this purpose, we adopt the conditional operator \Rightarrow_o , where expressions of the form $A \Rightarrow_o B$ are read "for organisation o , A counts as B ", adapting the idea presented and formally characterised in (Jones et al., 1996).

This conditional operator, allows us to represent our "transmission of agency" principles as follows:

$$\text{(trans.resp)} \quad (G_iOG_jA \wedge G_jA) \Rightarrow_o G_iA$$

$$\text{(trans.who.cares)} \quad (OG_jA \wedge A) \Rightarrow_o G_jA$$

Moreover, we may also deal with the representation of abstract tasks and the way these tasks are assumed to be decomposed within a given organisation o . An expression of the form $((A \wedge B) \Rightarrow_o C)$ - meaning that " $A \wedge B$ " counts for o as a way of achieving C - would represent a possible decomposition of C , and an expression of the form $(A \Rightarrow_o C) \wedge (B \Rightarrow_o C)$ would represent two alternative ways to get C .

The logic of \Rightarrow_o is that of a conditional logic of type CE (according to Chellas classification)⁴, with the additional axiom schemas:

$$(CC) \quad ((A \Rightarrow_o B) \wedge (A \Rightarrow_o C)) \rightarrow (A \Rightarrow_o (B \wedge C))$$

$$(CA) \quad ((A \Rightarrow_o B) \wedge (C \Rightarrow_o B)) \rightarrow ((A \vee C) \Rightarrow_o B)$$

We refer the reader to (Jones et al., 1996) for a detailed discussion of this logic, and for an account of its semantical model theory.

Different organisations may have different, indeed even conflicting, rules about what a given action means, or what it “counts as”. But it is clear that, given the truth of a sentence of type $A \Rightarrow_o B$, *then as far as organisation o is concerned*, if A is true then B is also true. Adapting for our present purposes the same general kind of strategy as is employed in (Jones et al., 1996), we introduce a relativised monadic operator, D_o , reading expressions of the form $D_o A$ as “ A is recognised by organisation o ”. We assign D_o a logic of type KD (Chellas, 1980), and we also adopt the following axiom schemas:

$$(\Rightarrow_o D) \quad A \Rightarrow_o B \rightarrow D_o(A \rightarrow B)$$

$$(\text{trans.meaning1}) \quad A \Rightarrow_o B \rightarrow D_o(E_i A \rightarrow E_i B)$$

$$(\text{trans.meaning2}) \quad A \Rightarrow_o B \rightarrow D_o(G_i A \rightarrow G_i B)$$

$(\Rightarrow_o D)$ says that if A counts as B within organisation o , then o recognises that if A is true then B is true. The schemas (trans.meaning1) and (trans.meaning2) reflect the fact that we also use the “counts as” relation to represent task decompositions.

Let us illustrate the additional inference capabilities provided by these notions. Consider the following example of an organisation o with four agents a, b, c and d , where the agents have just the capabilities indicated in Figure 2, where agent a has responsibility for q (that may be decomposed by $p1, p2$ and $p3$). We also assume that agent a has effective control over agent b with respect to q and b has effective control over agents c and d with respect to $p1$ and $p2$, respectively. We further assume that organisation o adopts the “transmission of agency” principle characterised by the schema (trans.resp) and that all the actions by the agents a, b, c, d are recognised by organisation o .

⁴ According to (Chellas, 1980), a conditional logic of type CE is a conditional logic closed under the rules:

$$(RCEA) \quad \text{if } \vdash A \leftrightarrow A' \text{ then } \vdash (A \Rightarrow B) \leftrightarrow (A' \Rightarrow B)$$

$$(RCEC) \quad \text{if } \vdash B \leftrightarrow B' \text{ then } \vdash (A \Rightarrow B) \leftrightarrow (A \Rightarrow B')$$

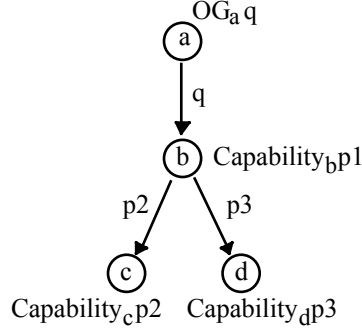


Figure 2.

Our logical system yields the following deduction:

$$\Gamma \vdash D_o(OG_{aq} \wedge G_{aq}),$$

where $\Gamma = \{((p1 \wedge p2 \wedge p3) \Rightarrow_o q), D_o(E_a OG_{bq} \wedge E_b OG_{cp2} \wedge E_b OG_{dp3} \wedge E_b p1 \wedge E_{cp2} \wedge E_{dp3}), D_o OG_{aq}\}$. Thus, in these circumstances, agent a as fulfilled his responsibility, as far as o is concerned.

Suppose now another organisation o' , and suppose that o' accepts the “transmission of agency” principle (trans.who.cares), i.e., $(OG_j A \wedge A) \Rightarrow_{o'} G_j A$. And consider the example described by Figure 2 in relation now to o' . Our logical system yields the following deduction:

$$\{((p1 \wedge p2 \wedge p3) \Rightarrow_{o'} q), D_{o'} OG_{aq}, D_{o'} (E_b p1 \wedge E_{cp2} \wedge E_{dp3})\} \vdash D_{o'} (OG_{aq} \wedge G_{aq})$$

This deduction is not forthcoming for organisation o . The examples show how the logic may be used to make explicit the fact that what counts for o' as a way for a to fulfil his responsibility does *not* so count for o .

These rather simple examples also suggest that these formal tools might be used in order to detect possible incompatibilities between different organisations in regard to the policies they adopt.

Finally, we should mention that the “counts as” relation might also be of value in capturing another idea commonly associated with *responsibility* in organisations: the idea that an agent, i , is *responsible to* some other agent, j . We encounter here another notion which has recently been discussed in connection with further refinement of the Kanger-Lindahl analysis of normative positions and rights: that of *directedness* of obligations, from the bearer of an obligation towards the agent to whom a duty is owed. (See, e.g., Herrestad et al., 1995.) In the organisational context, when i is responsible to j , this will often be understood to mean, at least in part, that i is answerable to j in the event of i 's failure to fulfil his responsibility;

this in turn implies, we suggest, that when *i* fails to meet his responsibility, *j* is *empowered* to initiate appropriate procedures against *i* (e.g., leading to the imposition of sanctions on *i*). It was just such a notion of *being empowered* that Jones and Sergot intended to capture in terms of the “counts as” relation (Jones et al., 1996).

6. CONCLUDING REMARKS

We have proposed a set of action concepts to describe aspects of organised interaction and we have illustrated their application to - in particular - the characterisation of transmission of agency and the analysis of the conditions under which a given organisation recognises that an agent has fulfilled his responsibilities. Some aspects of this formal analysis have been automated in a workbench capable of answering queries about what can and should be done, in a given organisation, to achieve particular goals (see Santos et al., 1996b).

Further analysis of the distinction between *direct* and *indirect* action, and of the range of notions associated with the concepts of influence and control, will be the subject of future work, together with further investigation of the applicability of the “counts as” relation. Our long-term aim is to help to supply logical underpinnings to the largely informal organisational modelling tools currently being employed in the systems sciences, where significant ambiguities in the central normative notions often appear to be left unanalysed.

Acknowledgements

This research has been carried out within the Portuguese research project n° PCSH/C/OGE/1038/95-MAGO, financially supported by JNICT, within the ESPRIT Basic Research Working Group 8319 MODELAGE (“A Common Formal Model of Cooperating Intelligent Agents”) and under the collaboration agreement between the Portuguese and Norwegian Research Councils.

References

- Belnap, N.: 1991, “Backwards and Forwards in the Modal Logic of Agency”, *Philosophy and Phenomenological Research*, vol. II, no. 4, 777-807.
- Brown, M. and Carmo, J.: 1996, “Report on the Third International Workshop on Deontic Logic in Computer Science”, *Knowledge Engineering Review*, vol. 11, no. 4.
- Chellas, B.J.: 1980, *Modal Logic - An Introduction*, Cambridge University Press.

- Elgesem, D.: 1993, *Action Theory and Modal Logic*, PhD thesis, Dept. of Philosophy, University of Oslo.
- Harel, D.: 1979, *First-Order Dynamic Logic*, LNCS 68, Springer.
- Harel, D.: 1984, "Dynamic Logic", in Gabbay, D.M. and Guentner, F., eds., *Handbook of Philosophical Logic II*, D. Reidel, Dordrecht, 497-604.
- Herrestad, H. and Krogh, C.: 1995, "Obligations Directed from Bearers to Counterparties", in McCarty, L.T. and Bench-Capon, T., eds., *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, ACM Press, New York.
- Hodge, B. and Anthony, W.: 1988, *Organization Theory*, Allyn and Bacon Inc.
- Hohfeld, W.: 1923, *Fundamental Legal Conceptions as Applied in Judicial Reasoning and Other Legal Essays*, W. Cook, ed., New Haven.
- Jones, A.J.I. and Sergot, M.: 1992, "Formal Specification of Security Requirements using the Theory of Normative Positions", *Proceedings of the European Symposium on Research in Computer Security (ESORICS'92)*, LNCS 648, Springer, 103-121.
- Jones, A.J.I. and Sergot, M.: 1993, "On the Characterisation of Law and Computer Systems: The Normative Systems Perspective", in Meyer, J.-J. and Wieringa, R., eds., *Deontic Logic in Computer Science: Normative System Specification*, John Wiley & Sons, 275-307.
- Jones, A.J.I. and Sergot, M.: 1996, "A Formal Characterisation of Institutionalised Power", *Journal of the IGPL*, vol. 4, no. 3, 427-443.
- Kanger, S.: 1957, *New Foundations for Ethical Theory*, Stockholm. (Reprinted in Hilpinen, R., ed.: 1971, *Deontic Logic: Introductory and Systematic Readings*, D.Reidel, 36-58).
- Kanger, S.: 1972, "Law and Logic", *Theoria* 38.
- Kanger, S. and Kanger, H.: 1966, "Rights and Parliamentarism", *Theoria* 32, 85-115.
- Lindahl, L.: 1977, *Position and Change - A Study in Law and Logic*, Synthese Library 112, D. Reidel, Dordrecht.
- Morris, P. and McDermid, J.: 1991, "Security and Normative Rights", in Meyer, J.-J. and Wieringa, R., eds., *Proceedings of the First International Workshop on Deontic Logic in Computer Science (Deon'91)*, Amsterdam, 294-314.
- Normatics: 1991, "Normatics: The Characterisation of Computer Systems and Complex Organisations as Normative Systems", *ESPRIT III BRA proposal*.
- Pörn, I.: 1970, *The Logic of Power*, Blackwell, Oxford.
- Pörn, I.: 1971, *Elements of Social Analysis*, Philosophical Studies, 10, Philosophical Society and Department of Philosophy, University of Uppsala.
- Pörn, I.: 1977, *Action Theory and Social Science: Some Formal Models*, Synthese Library 120, D. Reidel, Dordrecht.

- Pörn, I.: 1989, "On the Nature of Social Order", in Fenstad, J.E. et al., eds., *Logic Methodology, and Philosophy of Science VIII*, Elsevier, 553-567.
- Robbins, S.: 1987, *Organization Theory: Structure, Design, and Applications*, Prentice-Hall, New Jersey.
- Santos, F.: in preparation, "Organisation Theory - a Critical Overview".
- Santos, F. and Carmo, J.: 1996a, "Indirect Action, Influence and Responsibility", in Brown, M. and Carmo, J., eds., *Deontic Logic, Agency and Normative Systems*, Springer, Workshops in Computing Series, 194-215.
- Santos, F. and Carmo, J.: 1996b, "A Modal Action Logic Based Framework for Organization Specification and Analysis", in Alferes, J.J., Pereira, L.M. and Orłowska, E., eds., *Proceedings of the Fifth European Workshop on Logics in AI (JELIA '96)*, Springer LNAI 1126, 119-133.
- Santos, F., Jones, A.J.I. and Carmo, J.: 1997, "Action Concepts for Describing Organised Interaction", in Sprague, R.H., ed., *Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences (HICSS-30)*, vol. 5, IEEE Computer Society Press, 373-382.
- Scott, W. and Meyer J., eds.: 1994, *Institutional Environments and Organizations - Structural Complexity and Individualism*, Sage Publications.
- Tosy, H.: 1984, *Theories of Organization*, John Wiley & Sons.