Collective Agency, Obligations, Roles and Power to Act on Behalf of

José Carmo

Centre of Exact Sciences and Engineering, University of Madeira,
Campus Universitário da Penteada, 9020-105 Funchal, Madeira, Portugal

jcc@uma.pt

Abstract. A group of agents may cooperate and act jointly in order to achieve some common goals. But, if such common interests are not merely occasional, the group will organize its activities in a more permanent basis, creating an entity with a proper identity. These entities are intended to act in the world, but they cannot act directly: someone needs to act on their behalf. And such entities may be the subject of obligations and be responsible for their non-fulfilment (even legally, as for most organizations).

To achieve this, organizations are structured in terms of positions or roles, and the statute of the organization distributes its duties among the different positions, and attributes the power to act on behalf of the organization to the holders of some roles (when acting in such role). And these, through their acts, can create new obligations for the organization, e.g. by establishing contracts with other agents.

Concepts like role, and acting in a role, and normative concepts like obligation, permission, delegation, among others, constitute the basic building blocks in terms of which organizations are described, and our work has been to try to characterize such fundamental concepts through the use of modal logics. This paper follows this direction.

Keywords: agency, obligation, organizations, power to act on behalf of, role.

1 Introduction

Agents do act in order to achieve their goals. If an agent *a* has the ability[1] to achieve some result *p*, whenever he/she has such goal, we can say that *a* has the *power-of p*. According to Castelfranchi [10], this personal power[2] - "*power-of*" - is the basic notion of power.

[1] Like Castelfranchi (in [10]), we are here using ability in a broad sense, including not only physical ability, but also mental skills and the necessary resources.

[2] Castelfranchi distinguishes the *power-of doing some action* α, and the *power-of achieving some goal p* (by performing some action that leads to such result). Here

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

On the other hand, if an agent has some goal *p* and he/she lacks[3] the *power-of p*, the agent becomes dependent of other agents that have this power. When an agent *a* is dependent of an agent *b* for obtaining some goal *p*, and *b* knows this, *b* has *power-over a* with respect to *p* [10, pp. 222]. And in this way we enter into the level of the[4] *social powers*, relating (in this case) two agents.

We may have also the case where none of the agents of a group (set of agents) X, personally, has the power to achieve a certain goal *p*, but the members of the group, together (acting jointly), have such power. In such case we have (using Castelfranchi's terminology) a *collective power-of p* (or *co-power-of p*).

But, if the need to achieve some common goals is not merely occasional, the group will try to organize its activities in a more permanent basis, creating an entity with a proper identity, in order to act in the world as an agent. Let us call such entities of *collective agents*. These collective entities may be more informal, or more formal, recognized by the society and by the Law, as is the case of most *organizations*.

These entities have some features that make them very special and interesting. On one hand, they are like mini-societies, internally composed of a set of interacting agents, whose behaviour is typically regulated by norms. The norms define the expected behaviour of the agents (specifying the obligations, rights, responsibilities, or other possible normative concepts, that apply to them), but without excluding the possibility of deviation from the ideal behaviour, since we are in presence of autonomous agents.

On the other hand, externally, these entities will act within the world as if they were like any other agent (creating dependence relationships, making contracts, etc.). But they are special agents, at least in some respects. For instance, in general such collective agents do not act through a joint act of all his members, but through the acts of someone that acts on his name. But this means that there must exist norms that describe who has the *power to act on behalf of* the group, and in what conditions, and that such acts must be recognized and accepted as acts on behalf of the collective agent by the agents that interact with him and by the society in general. Note that this *institutional power* (as is called by Castelfranchi), of acting on behalf of the institution/organization, should not be confused with the *permission* of performing

we are thinking more in the latter, abstracting from specific actions (and we may see the former case as a particular case of the other, seeing *done(α)* as a goal).

[3] This may happen because the agent does not have the necessary resources, or skills, or simply because the agent does not believe that he/she has such power-of (he/she may have such power, but not be aware of this).

[4] In [10], Castelfranchi discusses four forms of social power: *comparative value*; *power-over* (and its 'species' of *incentive-power* and *rewarding-power*); *influencing power* (and its 'species' of *"command" power*); and *negotiation power.*

an act of such kind, neither with the *practical possibility* of doing it: see [21] for a discussion on this topic.

We also note that the representatives of the organization (collective agent) can make contracts on his behalf, establishing new obligations for the organization. Naturally, the organization may fail to fulfil some obligation and, when this happens, surely that (usually) it will not be the case that all the members of the organization will be responsible by such non-fulfilment. Thus it must be known how the organization will fulfil such obligations and who can be made responsible in the case of non-fulfilment.

What happens is that organizations are structured in terms of positions or roles, and the statute of the organization distributes its duties among the different positions, making their holders responsible for the fulfilment of such duties and, to that end, also attributing to them the power to act externally on behalf of the organization. And in this way we have a dynamic of obligations, where the obligations flow from the organization to the holders of some roles, and these, through their acts, may create new obligations for the organization.

At the same time, the organization's statute also gives to the holders of those positions the power to make a similar internal distribution of duties to the other members of the organization.

Typically there is a position - usually called president or director - with only one holder, that is the *leader* of the organization and that, by default, represents externally the organization. His holder not only has a great power over the other members of the organization (a "command-power", regulated by the statutes), as, given that he represents the organization, has also a increased power of influencing those that interact with him.

In organizations, the leaders are typically elected according to some rules (described in the statutes), and are chosen taking into account the influence they have on the voters and their capacity to convince them that they have a strategy that will benefit the organization. In less formal groups of agents (and less structured than organizations), the leaders may appear more informally, more spontaneously, but based on the same ideas: the capacity of influence over the other members of the group and of convincing them that they have a strategy that will benefit them, and that they can trust on them.

Our work has been to try to understand this complex issue of collective agency, and try to characterize some of the concepts that we consider essential for the representation of organized interaction.

For such characterization, we have been using modal logics, following the approach of Kanger, Pörn and Lindahl (among others). Our approach has been the following: we take advantage of previous contributions of applied modal logic to the representation of organized interaction; we confront their expressive power with further concepts relevant to the specification of organizations and organized collective agency in general; and, where necessary, we propose additional modalities in order to cope with those concepts.

We think that a combination of different kinds of modal logics will allow us to characterize those concepts that are the basic building blocks in terms of which an organization's policy is described, at an abstract appropriate level[5]. Given that we want to express norms and refer to actions, it is only to be expected that we need to use deontic and action logics. And conditional modal logics have been proposed to express the "count-as" relationships that seems to be essential to describe the institutional power referred above. There are other kinds of modal logics that are also useful for describing agents acting and interacting (for instance, as we have seen, belief and knowledge operators are also necessary for[6] the complete representation of some notions of power), but here we want to concentrate on the previous logics. The work herein presented should be seen only as a sketch of the possibilities open by such combinations (summarizing what has been our approach to these issues).

The rest of this paper is structured as follows. In section 2 we will make a very brief overview of the counts-as, deontic and action logics that we want to consider. In section 3 we will address the topic of (organized) collective agency, introduce the notion of role and of acting in a role, and show how those modal operators can be used to represent the associated relevant concepts. Conclusions will appear in section 4.

With respect to the notation, we will use: the Capital Latin letters D, E, F, H, O, P for modal operators; *a, b, ...* for (names of) agents; φ, ψ, ... for formulas (assertions about the states of affairs, etc.); *s* for the normative/legal system relevant (or for the "society"); X, Y, ... for groups of agents; *o* for organizations and collective agents; *r* for roles; and (in the semantics) w, v, ... for worlds (states). For the propositional connectives we will use: $\neg$, $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$; $\top$ will denote a tautology and $\bot$ a contradiction. The following precedence rules are assumed: 1st) unary operators; 2nd) $\wedge$; 3rd) $\vee$; 4th) the other binary operators.

2 A Brief Overview of Some Kinds of Modal Logics

In this section we will start with a brief presentation of the basic modal logic, and then we will make a *tour* around the variants (of modal logic) that we intend to use.

[5] We are here thinking on a first level of specification of an organization where the behaviour of the agents is described in terms of the abstract states of affairs that they bring about, and not necessarily immediately through the explicit reference to the concrete tasks they must perform (so as not to be immersed in too much details). Naturally, this first level of specification needs to be later refined, if we aim to construct complete models of organizations.

[6] And for the representation of other aspects of agent's modelling that are not of our specific concern here.

Although for the characterization of some aspects of organizations' modelling we will need first order modal logics (like those in [9] and [27]), to simplify the presentation, in this section (in this brief overview) we just consider a propositional setting. Thus, the atomic assertions of our formal language will be mere propositional symbols (e.g. p_1, p_2, ...), and the assertions (sentences) of the formal language (also called formulas) are built from the propositional symbols using the propositional connectives referred above and the modal operators.

2.1 Modal Logic

The basic modal language[7] just includes two modal operators: a necessity operator, usually denoted[8] by $\Box$, and a possibility operator, usually denoted by $\Diamond$. In fact, once the two operators are dual, we just need to consider one of them as primitive. For instance, we can consider $\Box$ as primitive, and define: $\Diamond = \neg\Box\neg$.

The *standard semantics* considers models of the form M=(W,R,V), where W is a non-empty set that denotes the set of possible worlds (or states of affairs), R is a binary relation on W, called the accessibility relation between worlds, and V is a valuation function that indicates which propositional symbols are assumed as representing true assertions at each world (for instance, we may consider that V maps each propositional symbol p_k into a subset of W, that is seen as the set of worlds where p_k is true).

We evaluate the truth-value of a sentence with respect to a world w of a model M. We write $M|=_W\varphi$ meaning that "the formula φ is *true at* the world w of the model M" (and we use $|\neq$ to deny $|=$), and define this notion recursively. The truth-value at a world w of the propositional symbols is defined by V: $M\ |=_W p_k$ iff (if and only if) $w\in V(p_k)$. The truth-value of the sentences built from the propositional connectives is evaluated, as expected, without changing the index w (i.e. without leaving the current world w in consideration) - e.g. $M|=_W\neg\varphi$ iff $M|\neq_W\varphi$; $M|=_W\varphi\wedge\psi$ iff $M|=_W\varphi$ and $M|=_W\psi$; $M|=_W\varphi\vee\psi$ iff $M|=_W\varphi$ or $M|=_W\psi$; etc. Finally, R is used to analyse the truth-value of the formulas built from the modal operators: φ is necessary at w iff φ is true at all worlds accessible from (or conceivable at) our current world w, i.e. $M|=_W\Box\varphi$ iff $M|=_V\varphi$ for every v such that[9] wRv (and, from $\Diamond=\neg\Box\neg$, it follows that $M|=_W\Diamond\varphi$ iff there exists some v such that wRv and $M|=_V\varphi$).

[7] For an introduction to modal logic, see e.g. [12].

[8] As we will see, when we consider other interpretations of these operators, we use usually other symbols for them, in order to suggest such interpretations.

[9] As usual, we write wRv instead of writing $(w,v)\in R$.

And we say that a formula φ is *true in* a model M, written M $|= \varphi$, iff it is true at all worlds w of M, and that a formula φ is *valid*, written $|= \varphi$, iff it is true in all the models we are considering.

We may, or not, impose some properties on the relation R considered in our models. Varying the properties we impose on R will vary (in general) the set of valid formulas. The logics (sets of valid formulas) we obtain in this way correspond to the *normal modal logics* (or *normal systems of modal logic*).

We can also give axiomatic characterizations for such logics, where the relevant formulas are then usually called theorems and are defined as the smaller set of formulas that can be obtained from a set of formulas, called axioms, by the application of some inference rules. Usually, it is used $|\text{-}\varphi$ to denote that φ is a theorem.

The minimal normal logic (that corresponds to the set of valid formulas that we obtain without imposing any property on the accessibility relation R) is called K, and can be defined[10] axiomatically as the smallest set of formulas that contains (as axioms) all tautologies and all instances of its distinctive schema[11]

(K) $\square(\varphi \to \psi) \to (\square\varphi \to \square\psi)$

and that is closed under the following inference rules, called Modus Ponens (MP) and necessitation rule (RN):

(MP) From φ and $\varphi \to \psi$ infer ψ

(RN) From φ infer $\square\varphi$

(i.e. "if $|\text{-}\ \varphi$ and $|\text{-}\ \varphi\to\psi$, then $|\text{-}\ \psi$" (MP) and "if $|\text{-}\ \varphi$ then $|\text{-}\ \square\varphi$" (RN)).

Stronger logics (with more theorems) can be obtained considering more axioms. For instance, looking at modal logic as the logic of necessity and pos-

[10] There are various ways of defining axiomatically these logics. It is possible to change the axioms and/or the inference (or deduction) rules, obtaining the same set of theorems. Although the modal logics we will consider may differ on their axioms and rules related with the modal operators, in what follows we will assume that they always incorporate the so-called *propositional calculus* (meaning that they contain as axioms, or at least as theorems, all instances of the tautologies, and that they satisfy the Modus Ponens inference rule, i.e. the set of their theorems is closed under the application of such rule).

[11] The acronym K was given in honour of the work in this area of Saul Kripke. We note that the same symbol is used to designate this logic and its distinctive axiom. Following Chellas [12], there is a practice of designating the modal logics by the sequence of the acronyms of their distinctive axioms (writing e.g. K, KT, KD, K4, KD4, etc.), although sometimes there are also other standard designations for some of these logics. For instance: KD is also simply designated as D; KT is simply designated by T; KT4 is also known as S4; and the well-known normal modal system S5 is nothing more than the logic KT5 (or the logic KT4B, that coincides with KT5).

sibility, it is natural to consider (at least) that "what is necessarily true, is true", which leads to impose also the axiom (schema)

(T) $\Box\varphi \rightarrow \varphi$

The normal modal logic so obtained (designated by KT or simply T: see the previous footnote) characterizes (has as theorems) the set of formulas that are valid in the class of models where R is reflexive (i.e., for any world w, wRw).

Sometimes we want to consider not logics stronger than K, but weaker (with fewer theorems than any normal modal logic). In particular, as we shall see, for some interpretations of the necessity operator, we do not want to have the following rule

(RM) If $\vdash \varphi \rightarrow \psi$ then $\vdash \Box\varphi \rightarrow \Box\psi$

known as the rule of the *closure of the necessity operator under logical consequence*, and a rule that, as can be easily seen[12], is derivable in any normal modal logic.

One way of defining the semantics of such non-normal modal logics is to consider the *minimal models* popularized by Chellas in [12] (also called of "neighbourhood semantics"). A *variant of such models*, that we have also considered, can be obtained replacing, in the standard models, the accessibility relation (R) by a function $f_\Box:2^W \rightarrow 2^W$, where $f_\Box(Z)$ is intuitively interpreted as denoting the set of worlds where the proposition[13] Z is necessary, and by defining the truth of $\Box\varphi$, at a world w of a model $M=(W,f_\Box,V)$, simply as follows:

$M \models_W \Box\varphi$ iff $w \in f_\Box(||\varphi||_M)$

(where[14] $||\varphi|||_M$ denotes the truth set of φ, i.e.$\{w: M \models_W \varphi\}$).

The logics that we can obtain through this semantics are called in [12] of[15] *classical modal logics* (or *classical systems of modal logic*). As before, we can get different logics by imposing different properties on our models, namely (now) on the function $f_\Box$. If we do not impose any condition, we get the smallest classical modal logic, called E, that can be defined as the smallest set of formulas that contains all tautologies and that is closed under Modus Ponens and under the *replacement of logical equivalents rule* (RE), below:

(RE) If $\vdash \varphi \leftrightarrow \psi$ then $\vdash \Box\varphi \leftrightarrow \Box\psi$

[12] From $\vdash \varphi\rightarrow\psi$, by the necessitation rule, we have $\vdash \Box(\varphi\rightarrow\psi)$. And $\vdash \Box\varphi\rightarrow\Box\psi$ follows, by Modus Ponens, from such theorem and from the K-axiom $\Box(\varphi\rightarrow\psi)\rightarrow(\Box\varphi\rightarrow\Box\psi)$.

[13] Taking propositions to be sets of possible worlds (namely the set of those worlds in which they are true).

[14] When the model M we are referring to is clear from the context, we write simply $||\varphi|||$ (instead of $||\varphi|||_M$).

[15] Note that the normal modal logics are a particular case of the classical modal logics (see [12]).

2.2 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 O. The dual of O ($\neg O \neg$) is denoted usually by[16] P and interpreted as meaning permission; and the concept of forbiddance is expressed through an operator F, that is defined as $O\neg$.

Standard deontic logic (SDL for short) sees O as a normal modal necessity operator. However, within this interpretation of the necessity operator, we do not want to have the T-schema ($O\varphi \rightarrow \varphi$) as a theorem of the logic (since what is obligatory may be violated / non-fulfilled), and SDL replaces this axiom (schema) by the weaker[17]

(D) $\quad O\varphi \rightarrow P\varphi$

That is, SDL corresponds to the smallest normal system of modal logic containing the D-schema.

Semantically, standard deontic logic considers serial standard models, i.e. models M=(W,R,V) where the relation R is serial (meaning that for each world w there is a world v such that wRv). The accessibility relation is then interpreted as follows:

wRv means that world v is an *ideal* version of world w

and, so, the definition

$M \models_W O\varphi$ iff $\forall_V$ (if wRv then $M \models_V \varphi$)

informally means that the formula φ is considered obligatory in a world w iff φ is true at all ideal versions of w.

SDL has been strongly criticized as an adequate system for deontic logic. Some of the critics that are made are referred next.

The D-schema corresponds to state that what is obligatory is permitted, which is intuitively a desired property; however, the D-schema is equivalent to $\neg(O\varphi \wedge O\neg\varphi)$, making it apparently impossible to express contradictory obligations in SDL. Also, once O is considered a normal operator (in SDL), by the O-necessitation rule, any tautology (more generally, any theorem) is obligatory, which is incompatible with the idea that obligations should be possible to fulfil and possible to violate. And SDL gives rise to a set of paradoxes, part of them related with the closure of the O-operator under logical consequence:

(RM)-rule: If $\vdash \varphi \rightarrow \psi$ then $\vdash O\varphi \rightarrow O\psi$

In order to solve these and other problems of SDL, many other deontic logics have been proposed. Although we do not want to enter in details, since the

[16] Some works use P to denote a primitive modal operator, intended to represent a concept of *explicit permission (*or *strong permission)*, seeing the dual of the obligation operator ($\neg O \neg$) as denoting (only) a (different) notion of *weak permission*.

[17] Weaker in the sense that it can be derived as a theorem in the logic T=KT.

deontic component, in isolation, is not our main focus here, we can briefly refer[18] that, among others, we may find proposals that define a semantics that:

- combine two accessibility relations (an ideal and a sub-ideal);
- consider ideal and sub-ideal worlds/states, plus ideal and sub-ideal transitions between worlds/states;
- consider the minimal models of Chellas [12], or some of its variants;

as well as proposals that consider dyadic modal operators, and a temporal dimension, or a preference ordering of the worlds, or contexts.

Besides the previous proposals, where the deontic operators apply to formulas, we also find proposals where the deontic operators apply to action terms, defining deontic logic on the top of a dynamic logic[19].

2.3 Action Logics: the 'Brings it About' Action/Agency Logics

There are various types of action logics. Here we concentrate on the logics of the "brings it about" type.

Contrarily to the dynamic logic operator (see the previous footnote), that is centred on the actions, the "brings it about" action operator is centred on the results, abstracting from the particular actions performed, which seems to be in accordance with the level of abstraction we need when we want to characterize and describe humans' acting, social interaction and complex normative concepts.

As it is stated in [35, page 4], "when dealing with the complexity of human affairs, it is sometimes difficult, or impossible, to pinpoint exactly what are the actions of an agent *x* by means of which some state of affairs *A* is brought about. This is especially true when there are actions of other agents to be taken into account ...". Furthermore, in some circumstances, we may even do not care which are such actions/means: sometimes what may be relevant is only that *a* has brought it about some state of affairs φ, and he is responsible for that; or that *a* is obliged to bring it about that φ, being possible that *a* can do it by different means.

This has lead to the development of logics of action, where actions are taken to be relationships between agents and the states of affairs that they bring about. The formal-logical development of this approach is due to Kanger and

[18] See e.g. [8], [17] and [26] for overviews of the paradoxes and of some of the proposals alternative to SDL.

[19] Dynamic logic was developed within Computer Science, related with the correctness proofs of programs (see [15] for an overview). It uses normal modal operators of the form [α], for α a program, where [α]φ means that "(if α is executed) after α, φ is the case". Later, these operators were applied to other kinds of action terms, and used generically to express the effects of actions.

Pörn [22], [28-29]. They introduce an action operator[20] (E) that relates an agent (a) with the effects of his action (φ), omitting details about the specific action that was performed and setting aside temporal aspects. The expression $E_a\varphi$ can be read as: "agent a brings it about that φ"; or "agent a sees to it that φ is the case"; or "agent a is responsible for its being the case that φ". Central to the "brings it about" concept is the notion of agency and of causation and responsibility.

A debatable question is if when we assert $E_a\varphi$, we should assume that agent a has brought about φ with intention, or not. On the "stit theory" initiated by Nuel Belnap and Michael Perloff (see e.g. [2-3]), a theory that also corresponds to a very important trend in this approach to the logic of action, the action operators are intended to be free from any psychological content such as belief, desire, and intention. On the other hand, some authors, like Tuomela [36], defend that the "sees to it" operator should be used only to describe intentional agency and intentional action. Hilpinen (in [19]) observes: "The expression 'seeing to it that φ' usually characterises deliberate, intentional action. 'Bringing about that φ' does not have such connotation, and can be applied equally well to the unintentional as well as intentional (intended) consequences of one's actions, including highly improbable and accidental consequences." Herein, following Mark Brown (see [6, footnote 1]), we are not considering such distinction between the expressions 'brings it about' and 'sees to it', and we assume that $E_a\varphi$ does not necessarily express that the indicated outcome (φ) of the agent's action was intended.

Although the formal properties assigned to the action operator E_a may vary among the different authors, with the main exception of Brian Chellas ([11], [13]), that proposes for E_a a normal modal logic of type KT, E_a is usually considered a non-normal modal operator satisfying the rule

(RE): If $\vdash \varphi \leftrightarrow \psi$ then $\vdash E_a\varphi \leftrightarrow E_a\psi$

and including the schemas:

(T) $E_a\varphi \rightarrow \varphi$

(C) $(E_a\varphi \wedge E_a\psi) \rightarrow E_a(\varphi \wedge \psi)$

(No) $\neg E_a\top$

(i.e. a classical modal operator of type ETCNo, according to the classification of Chellas in [12]).

The action operators (of this kind), herein considered, are also assumed satisfy these logical principles. The T-schema captures the intuition that if

[20] The symbol used to denote the action operator varies among the different authors.

agent *a* brings it about that φ, then φ is indeed the case[21]. Schema (No)[22] is used to try to capture the concept of *agency* itself: when $E_a\varphi$ is the case, the state of affairs φ is, in some sense, caused by or the result of actions performed by agent *a* (the result of *a*'s choices, in the stit terminology); the truth of $E_a\varphi$ must imply that the actions of *a* were necessary to get the state of affairs φ; no agent can meaningfully bring about what is logically true, or more generally, what was unavoidable.

With respect to the semantic characterization of the E_a operator, we find different proposals[23]. Here we just refer to three of the main approaches.

- Pörn [29] uses standard models with two accessibility relations, associated to each agent *a*,[24] $R1_a$ and $R2_a$, where informally $R1_a$ relates each world w with the worlds v where *a* has acted as in w, and $R2_a$ relates w with the worlds v where the agent acted differently from the way he has acted in w, or in the words of Pörn:

  w $R1_a$ v iff everything which *a* brings it about in w is the case in v

  w $R2_a$ v iff not everything which *a* brings it about in w is the case in v

  and defines[25]

  $M|=_w E_a\varphi$ iff $\forall_v$ (if $wR1_av$ then $M|=_v\varphi$) and $\exists_v(wR2_av$ and $M|\neq_v\varphi)$

  in order to capture the idea that $E_a\varphi$ is true (in a world) iff it is necessary for something *a* does that φ, but for *a*'s action it might have been the case that ¬φ (the "negative" or "counterfactual" condition)[26].
- Variants of minimal models are considered in e.g.[27] [14] and [32].

[21] Another way of arguing for the necessity of this schema is to follow Chellas [11] saying that an agent *a* can be held responsible for its being the case that φ, only if it is the case that φ.

[22] Considering this schema, we cannot have the rule (RM); otherwise (once $|- \varphi\rightarrow\top$), we would obtain $|- \neg E_a\varphi$, for any formula φ.

[23] See [19] for an overview of some of the main semantical devices that have been used.

[24] Pörn assumes $R1_a$ reflexive and $R2_a$ serial and irreflexive (i.e. for any w it is not the case that $wR2_aw$).

[25] This is equivalent to define E_a as a Boolean combination of two normal modalities.

[26] The proposal of Pörn, in [28], did not contain this counterfactual condition, giving a normal modal logic for E_a. On the other hand, in Kanger's definition in [22] ($M|=_wE_a\varphi$ iff $\forall_v$(if $wR1_av$ then $M|=_v\varphi$) and $\forall_v$(if $wR2_av$ then $M|\neq_v\varphi$)), the counterfactual condition was too strong (as was pointed out in [29], by Pörn).

[27] In [32], the models M include, for each agent *a*, a function $f_a:2^W\rightarrow 2^W$, where $f_a(Z)$ intuitively denotes the set of worlds where agent *a* sees to it proposition Z (functions that are constrained in order to get the desired principles for the action

- And, in the “stit theory”, a temporal semantics (much more elaborated) is proposed for the action operator, according to which an agent *a* sees to it that φ if the present fact that φ is guaranteed by a prior choice of *a*, meaning that in some previous time (the choice point) the agent made a choice that guarantee the truth of φ at the present moment, and at such choice point it would be possible for the agent to make another choice that would not guarantee the current truth of φ (the negative condition) [28].

Using E_a we can express several different positions in which an agent *a* might be with respect to a certain state of affairs φ, such as $E_a\varphi$ (did), $E_a\neg\varphi$ (averted) and $\neg E_a\varphi\wedge\neg E_a\neg\varphi$ (remained passive), as well as notions of control of other agents, like $E_aE_b\varphi$ (made *b* do), $E_a\neg E_b\varphi$ (made *b* avoid), etc. And combining E_a with deontic operators (combination to be discussed in the next section), 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 was done e.g. by Lindahl in [24].

As a simple example of the kind of analysis that can be made with such combination of operators, suppose we want to clarify the policy of some organization with respect to secret information. Suppose the regulation[29] says: "only the president and his secretary may have access to classified information". Consider that: *read(x,y)* means that agent *x* reads information *y*, *pre* denotes the president, *sec* denotes his secretary, *oth* denotes any other person, and *ci* denotes "classified information”. Does the regulation means only that

$$\mathrm{P}E_{pre}read(pre,ci)\wedge \mathrm{P}E_{sec}read(sec,ci)\wedge\neg \mathrm{P}E_{oth}read(oth,\mathrm{ci})$$

or does it also mean that there exists an obligation, both on the president and on his secretary, to see to it that no one else has access to the classified information, as it is expressed by the formula

$$\mathrm{P}E_{pre}read(pre,ci)\wedge \mathrm{P}E_{sec}read(sec,ci)\wedge\neg \mathrm{P}E_{oth}read(oth,\mathrm{ci})\wedge$$
$$\mathrm{O}E_{pre}\neg E_{oth}read(oth,\mathrm{ci})\wedge \mathrm{O}E_{sec}\neg E_{oth}read(oth,\mathrm{ci})$$

or such obligation falls only on the secretary (for instance, because she is the one that is responsible for storing such documents), as in

operator), and $M|=_w E_a\varphi$ iff $w\in f_a(||\varphi||)$. The semantic framework in [14] is much more elaborated and complex.

[28] Besides this action operator, related with agent's past choices, there exist also proposals of stit action operators related with agent's present choices (called “deliberative stit” operators). For a presentation of the stit theory, its problems, and a discussion of various stit operators, see e.g. [7].

[29] This is a version of the example of the hospital regulation analysed by Jones and Sergot in [20]. For the study of the theory of normative positions, its generation and its usefulness, with a much more detailed analysis of various examples, see [20] and [34-35].

$$\mathrm{PE}_{pre}read(pre,ci)\wedge\mathrm{PE}_{sec}read(sec,ci)\wedge\neg\mathrm{PE}_{oth}read(oth,\mathrm{ci})\wedge$$
$$\mathrm{OE}_{sec}\neg\mathrm{E}_{oth}read(oth,\mathrm{ci})$$

or does it still mean anything more? Naturally, these logics do not provide an answer to the question, but they can help in finding it, by providing the means to discriminate different possible alternatives for the formal representation of such regulation.

2.4 Deontic Logic Again: Personal Deontic Operators

It is natural to try to use deontic operators to direct personal individual agency, and in general we associate obligations (permissions or prohibitions) to agents. As it is said in [5, section 1]: "Obligations, I am supposing, normally call for action. We do not ordinarily impute obligations to beings we do not view as agents. Nor do we ordinarily hold that an agent has an obligation about which nothing (not even an exercise of restraint) should be done."

Thus, we may consider that we want to capture sentences meaning something like:

"agent a ought to see to it that φ"

where the obligation is attached to an agent, and the obligation is referring to an act that he must perform.

Suppose, then, that we consider explicit personal deontic operators to express such sentences. Concretely, suppose that we consider personal deontic operators (where the deontic operators are indexed by an agent), with the following informal meaning:

$\mathrm{O}_a\varphi$: agent a ought to see to it that φ
(or a is under an obligation of bringing about φ)
$\mathrm{P}_a\varphi$: a is permitted to see to it that φ
$\mathrm{F}_a\varphi$: a is forbidden to see to it that φ

Then, an obvious question to put, is if such personal deontic operators need to be primitive, or can be defined using other operators. And a natural proposal, suggested by the examples above, is to define the previous personal deontic operators as the following iterations of impersonal deontic operators and (indexed) action operators:

$$\mathrm{O}_a\varphi = \mathrm{OE}_a\varphi$$
$$\mathrm{P}_a\varphi = \mathrm{PE}_a\varphi$$
$$\mathrm{F}_a\varphi = \mathrm{FE}_a\varphi$$

Although this proposal seems natural, it has some consequences, and some criticisms have been made against this option in e.g. [16] and [23]. [30].

[30] Reductions of impersonal obligations to personal obligations are discussed e.g. by Hilpinen (in [18]), and by Herrestad [16] and Krogh [23], who criticize the reduc-

First we lose the inter definability of the personal deontic operators. We still have that $P_a\varphi\leftrightarrow\neg F_a\varphi$ (as well as $O_a\varphi\rightarrow P_a\varphi$, if O satisfies the D-schema), but the schemas $F_a\varphi\leftrightarrow O_a\neg\varphi$ and $P_a\varphi\leftrightarrow\neg O_a\neg\varphi$ are no longer valid. (Note that to impose the obligation $O_a\neg\varphi=OE_a\neg\varphi$ is "stronger" than to simply impose $F_a\varphi=O\neg E_a\varphi$). We do not think that this is a problem.

The second criticism, that may be called the "problem of transmission of obligations", can be described as follows: since (by the T-schema) $\vdash E_aE_b\varphi \rightarrow E_b\varphi$, if O satisfies the RM-rule (as is the case if O is a normal operator, as in SDL), then (according to the abbreviations above) the schema $O_aE_b\varphi \rightarrow O_b\varphi$ becomes a theorem, which is clearly unacceptable (what does the obligations of *a* have to do with the obligations of *b* ?).

Another criticism, that may be called the "problem of exclusion of conflicts", can be described as follows: since (by the T-schema) $\vdash E_a\varphi\rightarrow\neg E_b\neg\varphi$, if O satisfies the RM-rule, then $\vdash OE_a\varphi\rightarrow O\neg E_b\neg\varphi$, i.e. according to the previous abbreviations $\vdash O_a\varphi\rightarrow F_b\neg\varphi$, becoming impossible to express conflicts of obligations[31], even between different agents.

Taking into account these criticisms, we have two options: either consider that the personal deontic operators need to be primitive, or to consider that they can be defined as above, through iterations of impersonal deontic operators and personal action operators, but considering also that the impersonal obligation operator O does not verify the RM-rule, as was defended in [4]. Without excluding that the first option may also have some advantages, herein we will consider that the meaning of the personal deontic operators can be (at least intuitively) characterized by the above iterations of impersonal deontic operators and personal action operators, and assume that the O operator does not verify the RM-rule, at least with all its generality (not excluding that it can satisfy some weaker versions of such rule).

But how to characterize deontic concepts, like obligations, when we want to use them to direct the agency of groups of agents, organizations, etc.? This will be discussed in section 3, extending the analysis we have made in [9].

tions proposed in [18]. (In [23] a double indexed O operator is also proposed, in order to capture notions like "an agent *a* has an obligation with respect to an agent *b* to bring about some state of affairs".) For the definition of personal obligations based on dynamic logic see e.g. [30], where the previous works are also discussed.

[31] But, to be precise, as Mark A. Brown has stressed to us, in order to derive (as a theorem) the exclusion of conflicts of obligations schema $O_a\varphi\rightarrow\neg O_b\neg\varphi$, we also need that the operator O verifies the D-schema.

2.5 Some Extensions and Refinements of the 'Brings it About' Logics

Some generalizations, extensions and refinements of the "brings it about" type of action operators have been proposed. Below we will refer only to two of them. Besides these two, we have proposed also a generalization of these action operators in order to capture the notion of acting in a role, but we will delay its presentation to the next section 3, where we will motivate its interest.

Sometimes we have found it useful to have a way to express not-necessarily successful intentional actions, and in [33] we have proposed an *attempt* operator, H_a, with the following meaning:

$H_a\varphi$ means that "agent *a* has attempted to bring about that φ"

As expected, H_a does not verify the T-schema, and in [33] it was considered that H_a behaves like a classical modal operator of type EC (i.e. satisfying the RE-rule, "if $\vdash\varphi\leftrightarrow\psi$ then $\vdash H_a\varphi\leftrightarrow H_a\psi$", and including the C-schema, $H_a\varphi\wedge H_a\psi\rightarrow H_a(\varphi\wedge\psi)$, as an axiom).

In [33] it was also proposed the following bridging principle between the two action operators:

$E_a\varphi \rightarrow H_a\varphi$

However, since it is natural to assume that an attempt presupposes intention[32], and since we have considered that E_a does not presupposes intention, we now think that we should not impose this schema, as an axiom of our logic.

On the other hand, (at least for some applications) it seems acceptable to consider the schema

$H_aE_b\varphi \rightarrow H_a\varphi$ (and also, in principle, $H_aH_b\varphi \rightarrow H_a\varphi$)

Naturally, we can combine both action operators and use that to try to express some relevant concepts and discriminate different acts, and control situations. For instance, we may say that an agent *a* has an effective or total control

[32] The attribution of intention to the operator "attempt" is another reason for rejecting the RM-rule (i.e. that "if $\vdash\varphi\rightarrow\psi$ then $\vdash H_a\varphi\rightarrow H_a\psi$"), since the agent may be not aware of all the logical consequences of his attempts (and some consequences may even be not desired/intended). If we also consider epistemic logics, with K_a representing the knowledge modal operator "agent *a knows* that", then, intuitively, it might make sense impose the schema $H_a\varphi\wedge K_a(\varphi\rightarrow\psi)\rightarrow H_a\psi$ (as an axiom or theorem). However, we note that if we consider a normal modal logic for K_a (thus, assuming that an agent is an ideal reasoner that knows every theorem), then the previous schema will lead to the RM-rule. Another hypothetical alternative is to only impose that $H_a\varphi\wedge B_a(\varphi\rightarrow\psi)\rightarrow B_aH_a\psi$, with B_a representing the doxastic modal operator "agent *a believes* that".

("power-of") with respect to the state of affairs φ, if $H_a\varphi \rightarrow E_a\varphi$ is always the case.[33] And we can distinguish the following possible situations:

- $H_a\varphi \wedge E_a\varphi$

  ("*a* attempted and succeed" / "*a* has brought about φ with intention");
- $\neg H_a\varphi \wedge E_a\varphi$

  ("*a* has brought about φ without intention");
- $H_a\varphi \wedge \neg E_a\varphi$

  ("*a* attempted to bring about that φ, but he didn't succeed"); and
- $\neg H_a\varphi \wedge \neg E_a\varphi$

  ("*a* has not brought it about that φ, neither has attempted")[34]

as well as we can use iterations of the action operators to try to discriminate situations[35] like:

- *a causes that b acts, that b performs an action that produces* φ

  (*a*'s action is the cause of *b*'s acting: either *a* motivates, persuades, induces *b* to act, or *a* simply provokes (by physical or reactive mechanisms) *b*'s behaviour. A possible example: *a* is the principal while *b* is the hired killer).

  Possible representation:

  $H_aE_b\varphi \wedge E_aE_b\varphi$

  (assuming that b has succeeded in bringing it about that φ)

  or (maybe better, possibly depending on the cases)

  $H_aH_b\varphi \wedge H_b\varphi \wedge E_b\varphi$

  (again, assuming that b has succeeded in bringing it about that φ)
- *a causes that b's independent action produces* φ

  (*a*'s action does not cause *b*'s action but only that *b*'s action produces the outcome φ. A possible example: *a* prepares a bomb connected with *b*'s room switch, and *b* ignores everything).

  Possible representation:

  $H_aE_b\varphi \wedge \neg H_b\varphi \wedge E_b\varphi$

[33] This should be seen only as a possible example, since the discussion of how to represent this control's notion falls outside the scope of this paper. For instance, as we have referred at the introduction, in order to have the real power, the agent must be aware of having such power, which means that probably we should add to the definition above that $B_a(H_a\varphi \rightarrow E_a\varphi)$ (with B_a representing "*a* believes that", as in the previous footnote).

[34] Note that the last two cases are both compatible with a situation where φ is the case (because someone else has brought about φ), as well as with a situation where φ is not the case.

[35] This problem was posed to us by Cristiano Castelfranchi, and what follows is an attempt of a possible solution.

On the other hand, sometimes, we have also found useful be able to distinguish between the direct and immediate effects of an agent's own actions and those indirect effects that follow from such direct effects, either sometime later, by some causal connection, or by institutional connection (a topic to be addressed later in this paper), or by other reasons, effects that can also be attributed to the responsibility of the agent.[36] Thus, we will consider a "*direct*" action operator[37], denoted by D_a, with $D_a\varphi$ meaning that "agent *a* has brought it about that φ is the case, directly (and immediately)", and we will continue to use the operator E_a to express states of affairs that are generically brought about by the agent *a*, in some generic sense.

As a natural bridging principle we have that

$$D_a\varphi \rightarrow E_a\varphi$$

but not the other way around. For instance, suppose that agent *a* steals the canteen of *b*, that was in the desert, and in consequence of this, some hours later, *b* dies of thirst (dehydrated): in such a situation, we may state that $E_a$dies(*b*), but not that $D_a$dies(*b*) (although it is correct to state that D_a"*b* is without water", as well as D_a"eventually *b* dies"). On the other hand, a situation where *a* shoots a gun, killing (immediately) *b*, will be correctly described by $D_a$dies(*b*).

In the applications we have in mind in this article, we are not especially interested in distinguishing the direct effects of an agent's actions and those indirect effects that follow, sometime later, by some causal connection (as in the previous example), but we will be particularly interested in distinguishing the direct effects of an agent's actions from the institutional, or legal, consequences that might follow from them, in some circumstances, to be discussed later.

2.6 Counts-as

Another modal operator that we want to use in order to characterize collective agency and organizations' modelling is the "counts-as" operator proposed by Andrew Jones and Marek Sergot.

According to Jones and Sergot [21], within institutions, organizations, or other normative systems, there are rules that state that some acts, or some

[36] In [32] we have also introduced distinct action operators to distinguish the cases where an agent brings it about a state of affairs φ, by himself, and the cases where the agent succeeds in obtaining φ "indirectly", by exercising his power or influence on other agent (that brings it about φ). Here, we want to describe not only what an agent has brought about by himself, but, more specifically, the direct and immediate effects of the agent's own actions.

[37] For a definition of this operator within the stit framework see [7].

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). And, to express such "*count as" ("meaning") relations* they propose the use of a conditional modal operator $\Rightarrow_s$ (where the index *s* refers the relevant system of analysis).

Although they do not restrict the use of $\Rightarrow_s$ connected to act-descriptions, this was their main goal, considering expressions like $E_a\varphi \Rightarrow_s E_b\psi$ with the following intended meaning: "*a*'s act of bringing about φ, *counts*, *within* the institution *s*, *as* a means by which agent *b* (who may but need not be the institution *s* itself) establishes state of affairs ψ; *a* may be said to act on behalf, or as an agent of *b*".

For $\Rightarrow_s$ they proposed a modal conditional logic containing (the tautologies, the Modus Ponens rule, and) the following two rules

If $\vdash \psi1 \leftrightarrow \psi2$ then $\vdash (\varphi \Rightarrow_s \psi1) \leftrightarrow (\varphi \Rightarrow_s \psi2)$

If $\vdash \varphi1 \leftrightarrow \varphi2$ then $\vdash (\varphi1 \Rightarrow_s \psi) \leftrightarrow (\varphi2 \Rightarrow_s \psi)$

plus the following schemas (but not their converses)

$((\varphi \Rightarrow_s \psi1) \wedge (\varphi \Rightarrow_s \psi2)) \rightarrow (\varphi \Rightarrow_s (\psi1 \wedge \psi2))$

$((\varphi1 \Rightarrow_s \psi) \wedge (\varphi2 \Rightarrow_s \psi)) \rightarrow ((\varphi1 \vee \varphi2) \Rightarrow_s \psi)$

Andrew Jones and Marek Sergot have also introduced a normal modal operator, of type KD, denoted by D_S, where expressions of the form $D_s\varphi$ may be read as follows: "it is a constraint of institution *s* that φ is the case" or "it is incompatible with the rules operating in institution *s* that φ is not the case". Herein we use D for direct action. Thus we will use R_s instead of D_s, and read $R_s\varphi$ as meaning "according to the rules operating/accepted in institution *s*, φ is the case" or "is *recognized*/accepted by *s* that φ is the case".

By the fact that an institution *s* considers that φ is the case (according to its internal rules), we cannot conclude that φ corresponds to a true fact of the world.[38] Moreover, what it is recognized/accepted by an institution may differ from institution to institution: for instance, some institution (e.g. a religious institution) may consider that, according to its rules, "John and Helen are married", while other institutions (e.g. the normative system of some country) may consider that, according to its rules, "John and Helen are not married".[39]

[38] For instance, someone may be innocent of a crime, although the normative system *s* (through its courts) may declare that (consider proved that, according to its rules) he is guilty.

[39] With respect to some facts, corresponding to normative (or institutional) relationships, like "John and Helen are married", it is even arguable if we can talk about its truth, or if we can only refer that they are recognized as being true by some institution *s*. In such cases what happens is that usually we identify its truth as meaning truth according to some relevant legal system. However, an analysis of this topic lies outside the scope of this paper.

Thus the operator R_S cannot verify the T-schema, $R_S\varphi \rightarrow \varphi$. And Andrew Jones and Marek Sergot also reject the converse of the T-schema: $\varphi \rightarrow R_S\varphi$.[40]

Finally, as bridging principles between $\Rightarrow_S$ and R_S, Jones and Sergot have proposed (in [21])

$$(\varphi \Rightarrow_S \psi) \rightarrow R_S(\varphi \rightarrow \psi)$$

(from which it follows $(\varphi \Rightarrow_S \psi) \rightarrow (R_S\varphi \rightarrow R_S\psi)$, since R_S is normal). Moreover, they also adopted the following (more debatable) schema

$$(\varphi \Rightarrow_S \psi) \rightarrow (\varphi \rightarrow R_S\varphi)$$

as a means of securing the stronger detachment principle:

$$(\varphi \Rightarrow_S \psi) \rightarrow (\varphi \rightarrow R_S\psi).$$

3. Organized Collective Agency

Agents may be subject of obligations and in order to fulfil them, they must act. And, by acting and interacting, the agents can modify the relevant state of affairs, as well as create new obligations (for instance, by making contracts). And, as we have already seen, we may combine impersonal deontic operators and personal action operators (for instance of the brings it about type) to describe the obligations that apply to each agent, and the effects of their acting on the relevant state of affairs. But, if we want to describe social interaction, we cannot avoid considering joint actions and collective agency.

3.1 Joint Action

Two or more agents can jointly act in order to do some task (e.g. to move a very heavy table, to make a contract, etc.), and the brings it about action operators can be generalized in order to cover also such situations, as was proposed for instance by Lars Lindahl in [24].

Supposing that X denotes a (finite) set of agents, we may write $E_X\varphi$ with the following informal meaning: the set of agents described in X jointly see to it the state of affairs φ. In general, when we assert $E_X\varphi$ we want to express that the actions of the agents in X cause the state of affairs φ; the actions of

[40] Once R_S is a normal operator, (if φ is a *logical truth*, in the sense of a theorem of our logic, i.e.), if $|\text{-}\varphi$, then $|\text{-}R_S\varphi$, and so, in such case, also $|\text{-}\varphi \rightarrow R_S\varphi$. However $\varphi \rightarrow R_S\varphi$ is not a theorem *schema*, i.e. is not a theorem for all formulas φ, and this is in accordance with the meaning of R_S. Note that even when φ represents what we may call a "scientific truth", its truth may be not recognized by all the institutions; e.g. for a long time the Catholic Church has not recognized that the earth moves around the sun.

each of such agents were necessary (or, at least, contribute) to the production of φ. We may say that the agents described in X jointly cooperate to bring about that φ 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[41] $E_{\{a,b\}}\varphi \wedge \neg E_a\varphi \wedge \neg E_b\varphi$ (with $a \neq b$) can be consistent, allowing to express situations where two agents jointly have brought about some state of affairs, without being the case that any of them has brought it about by himself (as when four grams of poison is the minimum quantity sufficient to kill a person *c*; *a* and *b*, each give to *c* two grams of poison at the same time; and φ is the sentence "*c* dies": [24, page 222]).

There exist also situations where we may have $E_{\{a,b\}}\varphi$ and $E_a\varphi \wedge E_b\varphi$ both true.[42] And there might exist even cases where the production of some state of affairs φ by an agent *a* "counts as" if it was the set of agents $\{a,b\}$ that have produced φ. However, by obvious reasons, we reject a general principle of the form

$$E_X\varphi \rightarrow E_Y\varphi \text{ , for } X \subseteq Y$$

Using this operator, we can describe the establishment of contracts. For instance, the establishment of a contract between *a* and *b*, by which *a* becomes under the obligation of doing φ and *b* becomes under the obligation of doing ψ, can be expressed by $E_{\{a,b\}}(O_a\varphi \wedge O_b\psi)$.

3.2 Collective Agency

Suppose now that a group (a set) X of agents wants to act collectively in a more or less permanent basis.

Then, one *first hypothesis* that we may consider is that "whenever the group X wants to act, all members of X meet and act together", and we could use the previous action operators to express such situations.

But, if we assume that the group X of agents wants to act collectively in a more or less permanent basis, probably it will interact with other agents and groups, making contracts, etc., creating in this way obligations for the group itself. And the question now is how we can characterize such kind of *collective obligations* (that below we will denote by) O_X.

[41] Lindahl would index the action operator with *a*+*b* instead of with $\{a,b\}$: see [24, pp. 220-222).

[42] According to [24, page 222], this happens in the example above, if both *a* and *b* give four grams of poison to *c*. In [7] we defend other point of view regarding this example.

Since we are assuming that the group X always acts through a joint act of all its members, it is natural to assume that an obligation $O_X\varphi$ will have the form of an obligation on a joint action of the group X, which we may represent by an expression like $OE_X\varphi$. However, *this does not solve our problem*. As a matter of fact, as we defended in [9], we think that only an agent acting can be deontically qualified, because if an obligation is not fulfilled we must know who is potentially subject to punishment. According to this point of view,[43] the collective obligation must be defined in terms of obligations of the agents, and, at a first sight, in terms of individual obligations of the members of X. Two options seem then apparently natural (where we follow the terminology of [31]):

1) A "general obligation", where an obligation on a group corresponds to an obligation on each of its members: $O_X\varphi = \forall_{x\in X}\, O_x\varphi$
2) Or a "weak general obligation", where an obligation on a group corresponds to an obligation on some of its members: $O_X\varphi = \exists_{x\in X}\, O_x\varphi$

Option 2) does not serve. To see this suffices to note that it validates $O_X\varphi \rightarrow O_Y\varphi$, for any Y such that $X \subseteq Y$.

In the case under analysis, where the group X always acts through a joint act of all its members, it seems natural to say that an obligation on the group corresponds to an obligation on each of its members. But we must be careful when describing the obligation that applies to each of the members of the group X. We do not want to say that $O_X\varphi$ means that $\forall_{x\in X} O_x\varphi$ (i.e. $\forall_{x\in X} OE_x\varphi$), as it would be stated according to option 1. Suppose, as a very simple example, that X is a football group/team and φ is "to score (at least) five goals on today's game"; then, in order that the team score five goals, we do not need that each of his players score five goals (and thus we do not want to say that each of his players is under that obligation). In this case (of a group X that always acts through a joint act of all its members), the most natural interpretation of $O_X\varphi$ seems to be $\forall_{x\in X} OE_x E_X\varphi$ (i.e. $\forall_{x\in X} O_x E_X\varphi$), or, using the attempt operator, $\forall_{x\in X}\, OH_x E_X\varphi$.

3.3 Collective Agency: Acting in the name of, Counts-as and Direct Acts

Let us continue to suppose that a group X of agents wants to act collectively in a more or less permanent basis. However, the previous case, where the group always acts through a joint act of all its members, is not the most usual case. If a group X wants to act in a more or less permanent basis, usually the

[43] Herein we will only consider obligations. See [9] for a similar discussion related with permissions, and prohibitions.

group will organize its activities in some *stable way*, and allowing that some acts may be performed *in the name of the group* by some of his members. In that case, the group will create a *statute* (or *an internal code*) stating that such is the case.

Suppose, for instance, that the group decides that:

i) His member a may bring about a certain type of states of affairs φ in the name of the group; or that

ii) Any of his members may bring about the state of affairs φ in the name of the group (suppose that X is a group of Mafia killers and they have a code stating that "when one of us kills, we all kill"[44]).

If we want to characterize such situation, what it will be required? One hypothesis would be to describe such decision-act of the group X as follows:

case i): $E_X (E_a\varphi \rightarrow E_X\varphi)$

case ii): $E_X \forall_{x\in X} (E_x\varphi \rightarrow E_X\varphi)$

However, these formal characterizations (of what the group X has brought about through its decision) have some imprecision and drawbacks.

Consider e.g. case ii), and suppose that a is a member of the group, $E_a\varphi$ is the case and φ means kills(b). From $E_X\forall_{x\in X}(E_x\varphi\rightarrow E_X\varphi)$ it follows that $\forall_{x\in X}(E_x\varphi\rightarrow E_X\varphi)$ and so (since $a\in$X) also $E_a\varphi\rightarrow E_X\varphi$. Thus, from $E_a\varphi$, we can derive that $E_X$kills(b). But, in such a situation is it correct to conclude that $E_X$kills(b) ?

Our first comment, regarding such situation, is that the sentence $\forall_{x\in X}(E_x\varphi\rightarrow E_X\varphi)$ may describe only an internal agreement of the members of X, not necessarily accepted by the "external world" (the "society", or possibly better, the "relevant normative/legal system", in what follows denoted by s). For instance, in the previous case the normative system may not recognize that $E_X$kills(b) is the case, if a has killed b alone[45].

[44] Note that this code is compatible with a further requirement that any member of the group is forbidden to kill without an express joint authorization of the group X, but, nevertheless, if someone of X kills (let us suppose) some member b of another rival group Y, the group X will assume that this counts as if the group X has killed b, although possible sanctions might then be taken by the group X with respect to the real killer of b.

[45] Observe that the situation described is not exactly like a situation where group X jointly decides that his member a has a duty to kill agent b, which we could represent by $E_X OE_a$kill(b). In such case, if a kills b, the normative system may recognize, or not, that both $E_a$kills(b) and $E_X$kills(b) are true in that situation.

Thus, in such case, we should replace the material implication operator (→) by the counts-as operator of Andrew Jones and Marek Sergot, and write (considering e.g. case ii)) that

$$E_X \forall_{x\in X} (E_x\varphi \Rightarrow_X E_X\varphi)$$

Then, assuming $a\in X$, we can deduce $E_a\varphi \Rightarrow_X E_X\varphi$. And if $E_a\varphi$ is the case, although we cannot derive that $E_X\varphi$, we can derive that $R_XE_X\varphi$ is the case, where an expression like $R_X\psi$ can be read as follows: "according to the rules accepted by X – by the members of X – ψ is the case" or "is recognised/accepted by X (interpreted as it is recognised/accepted by all the members of X) that ψ is the case".

Of course, there might exist cases where the legal system gives to X the power to allow someone act on his name with respect to some state of affairs φ (for instance, according to the normative system, by signing an appropriate document, a family X may give power to an agent *a* to sell the family's house[46]). In such cases, it seems that the act of exercising such power by the group X can be describe as follows (assuming that *a* is the name of the agent to whom X has delegated the power to bring about φ on its behalf):[47]

$$E_X (E_a\varphi \Rightarrow_s E_X\varphi)$$

[46] Situation that could be characterized as follows:

$$(*)\ E_X\text{document}(a,X) \Rightarrow_S (E_a\text{sells}(\text{house\_of}(X)) \Rightarrow_S E_X\text{sells}(\text{house\_of}(X)))$$

According to this characterization, if $E_X\text{document}(a,X)$ is the case (where $\text{document}(a,X)$ means an appropriate document is signed by X with respect to agent *a*), then, according to the logic proposed in [21], we can derive that

$$R_S(E_a\text{sells}(\text{house\_of}(X)) \Rightarrow_S E_X\text{sells}(\text{house\_of}(X))).$$

And, if we accept also the schema $(\varphi\Rightarrow_S\psi)\leftrightarrow R_S(\varphi\Rightarrow_S\psi)$, then we can also derive

$$E_a\text{sells}(\text{house\_of}(X))\Rightarrow_S E_X\text{sells}(\text{house\_of}(X)).$$

Moreover, if we assume the schema $(\varphi\Rightarrow_S\psi)\rightarrow(E_X\varphi\rightarrow E_XR_S\psi)$ (which needs further research, since it is not a schema of [21]), besides the schemas $E_X\varphi\leftrightarrow E_XE_X\varphi$ and $(\varphi\Rightarrow_S\psi)\leftrightarrow R_S(\varphi\Rightarrow_S\psi)$, then, from (*), if $E_X\text{document}(a,X)$ is the case, we can deduce

$$E_X(E_a\text{sells}(\text{house\_of}(X)) \Rightarrow_S E_X\text{sells}(\text{house\_of}(X))$$

which seems to correctly characterize the situation occurring.

The signing of the appropriate document is the means by which X has brought it about that $(E_a\text{sells}(\text{house\_of}(X))\Rightarrow_S E_X\text{sells}(\text{house\_of}(X))$ is the case.

[47] *When* we do not want to refer to different legal systems, *sometimes we identify the truth* (of some kind of statements) *with* its recognition by the "relevant legal system". In such cases, *for practical purposes, we may dispense the "counts-as" operator*, and write $E_X(E_a\varphi\rightarrow E_X\varphi)$ instead of $E_X(E_a\varphi\Rightarrow_s E_X\varphi)$.

However, there are still other reasons by which we think that neither of the formulas

$E_a\varphi \Rightarrow_X E_X\varphi$ (or, in case ii), $\forall_{x\in X} (E_x\varphi \Rightarrow_X E_X\varphi)$)

$E_a\varphi \Rightarrow_s E_X\varphi$

represents exactly the state of affairs that the group X has created (or wants to create).

In fact, not all acts of *a* are made in the name of the group X and it is possible that agent *a* may bring about φ not in the name of the group X, but for himself (or in the name of another agent or group), in which case we do not want derive $R_XE_X\varphi$ nor $R_sE_X\varphi$, from $E_a\varphi$. Even in the previous example of a Mafia group X with the code "when one of us kills, we all kill", it is implicitly assumed that it means "when one of us kills (acting as a member of the group), we all kill"; if a member of the group is paid by another organization to kill someone, such act will not certainly be assumed by the group X as an act made by the group.

An agent can do a similar act playing different roles, but to know the effects of such act and its deontic classification, we must know in which role it was played.

For instance, an administrator of a company may be permitted to drive a company's car when on duty – i.e. when he is acting in the quality of administrator – but be forbidden to use that car when on holiday; and even if he is permitted to drive that car on holiday, if he has a car accident, the responsibility of repairing the damage caused will depend on the role he was playing when he had the car accident (probably the company will be responsible for repairing the damage if, and only if, he was on duty). As another example, a person that is administrator of a computer system can interact with the computer system in the quality of administrator or as a simple user, and the effect of its orders (e.g. to delete certain type of files) may depend the role that he is playing (either the deletion, or nothing, because a simple user is not authorized to delete such files).

Thus it becomes necessary express the quality in which agent *a* has acted when he brought about φ. Using $E_{a:X}\varphi$ (or $E_{a \text{ as } X}\ \varphi$) to denote that *a* has brought about φ as if it was X that has acted (*a* has brought about φ acting in the name of X - as a representative of X), then we can write (where $E_{X:X}\varphi$ may be read "X has brought about φ acting as himself)

$E_{a:X}\varphi \Rightarrow_X E_{X:X}\varphi$

as well as

$E_{a:X}\varphi \Rightarrow_s E_{X:X}\varphi$

Naturally, we can question how we know that a has brought about φ as himself (in his own name) or in the name of the group X. We will discuss such issue later.

Another question is how we can discriminate, in our formal language, the cases where

i) X has brought about φ, because some agent has brought about φ in the name of X

from the cases where

ii) X has brought about φ, directly, by himself (by a joint act of the group X)

which might be important to know, e.g. for legal purposes (particularly if some illegal act has been made).

For instance, we can introduce a notation like

$E_{X\,/_s\,a}\,\varphi$ (or simply $E_{X/a}\,\varphi$, if the institution s can be deduced from the context)

(read as "according to the institution s, X has brought about φ through a's acting"), as an abbreviation of a statement like

$$E_{a:X}\varphi \wedge (E_{a:X}\varphi \Rightarrow_s E_{X:X}\varphi)$$

However, although this abbreviation is useful, we note that there might exist cases where $E_{X/_s a}\varphi$ is true, but X has also directly jointly brought that φ is the case[48]. Other option is to consider the operator of direct action, D, to describe the direct acts made by an agent and the direct joint acts made by a group X of agents, and discriminate i) and ii) above as follows:

i) $E_{X:X}\varphi \wedge \neg D_X\varphi$

and

ii) $D_X\varphi$

Finally, how should we characterize now a "*collective obligation*" $O_X\varphi$ for the kind of group X of agents under analysis?

Since we are not in presence of a group X that always acts through a joint act of all his members, allowing that some members can act on his name, we should not interpret $O_X\varphi$ as meaning $OD_X\varphi$. A reasonable interpretation might be $OE_{X:X}\varphi$. But how to guarantee that such obligation will be fulfilled, and how to know whom may be subject to punishment, if such is not the case?

[48] Suppose a hypothetical case where (a single agent or) a group X has hired a specialist a to kill (let us say) the president, but meanwhile all members of the group X have also put poison in the glass of water of the president, and at the same time the president drinks the glass and receives a shot from a.

Clearly, now this collective obligation $O_X\varphi$ should not be seen as a general obligation of the form

$$O_X\varphi = \forall_{x\in X}\ O_x\varphi$$

or of a similar kind.

Although the "weak general obligation" $O_X\varphi = \exists_{x\in X}\ O_x\varphi$ still does not is what we want (by the reasons already explained), in some sense it gives us a kind of "meta-rule", stating what is that the group X must do (when it creates his statute). Basically, the group must guarantee that someone becomes responsible by fulfil the obligations of the group:

$$O_X\varphi \rightarrow \exists_{x\in X}\ O_x\varphi$$

More precisely, the group will state in its statute, something like

$$O_X\varphi \rightarrow OE_{a:X}\varphi$$

where the particular agent *a* may be dependent on the type of statement φ to which the obligation refers (and the group will also give representative powers to such acts).

3.4 Collective Agency: Collective Agents and Roles

Suppose now that, as before, a group X of agents have common interests and want to act collectively in a more or less permanent basis, but that such common activity should continue, even if some member of the group is not anymore interested in it, or when it is possible to aggregate other members to the group.

In such cases, the natural way for the group X to proceed is to create a distinct entity, with its *own identity* (like it is the typical case of an organization). The members of the group will be related with such entity by special relationships, like "member-of" ("associate-of", etc.), but such entity will persist even if the set of its members will change.

Of course, this entity, created by the group X, needs to act, and so it is an agent. In [9] and [27] we have called it of "institutionalized agent"[49]; herein we will call it a "collective agent" (independently of the number of persons related with it). In some cases (like when we talk about organizations), the Law will recognize such entity as a "real agent" (sometimes called an "artificial person"), having juridical personality and legal competence, as any natural person. In other cases, that we do not want to exclude here, this "collective agent" may be more informal, without a legal recognized status.

Naturally, this "collective agent" needs to act, but *it cannot act directly*! *Thus, someone needs act on his name*. When a "collective agent" *o* is created,

[49] Possibly, an alternative, better name, is *institutional agent*.

a statute is elaborated, defining the main norms regulating *o*'s activity, and stating, in particular, the rules by which one can act on his behalf.

The main difference for the previous case is that such rules do not state who is the concrete person that can act in the name of the collective agent.

The "collective agent" (the organization) is usually structured in terms of what we may call *positions*, or *roles* within the organization (we may call them *structural roles*[50], meaning that they correspond to roles defined in the structure of the organization), and the statute of the organization describes who has the power to act in the name of the organization. But this description is abstract: it does not say which particular person can act in the name of the organization; instead, it attributes such power to the holders of some positions/roles (independently of whom they are).

Normally, exists an individual position (in the sense that may have only one holder), like *president-of*, whose titular can act on behalf of the organization and is usually seen as the *leader* of the organization, and, when the organization is created, the statute not only defines the rules for his election/choice, but it also defines a provisory initial titular for that position. The statute also defines if the titular of such position has power to delegate the power of acting in the name of the organization, and in which conditions.

Leaders are typically chosen based on their influence over the other members of the group and on their capacity of convincing them that they have a strategy that will benefit the organization. It is important for the success of the organization that their members trust on their leader, but we will not discuss such issue here.

As we have referred, to some roles are attributed representative powers. Suppose that r:REP(o,φ) means that "the role r is a representative role of the collective agent (organization) o, with respect to a state of affairs φ (the scope of the representation)". The notation r:REP(o,φ) doesn't mean that role r can act in the name of o. We think that *(only) agents can act, and roles are not agents* (thus a role does not act). What r:REP(o,φ) does mean is that when someone, playing the role r, does (brings it about that) φ, this act can be seen as an act made in the name of o (as if it was o who has acted), which was expressed in [27] by

$$\forall_x \, (\mathrm{E}_{x:r}\varphi \rightarrow \mathrm{E}_{o:o}\varphi)$$

where $\mathrm{E}_{x:r}\varphi$ means "x acting in the role r does (brings it about that) φ" and $\mathrm{E}_{o:o}\varphi$ means "o acting in the role of itself does φ".

Using the counts-as operator, we can reformulate the previous definition as follows, assuming that this representative power is recognized by the "society" s (the relevant normative system)

[50] Castelfranchi [10] refers to them as *positional entities*.

$$\forall_x\,(\mathrm{E}_{x:r}\varphi \Rightarrow_s \mathrm{E}_{o:o}\varphi)$$

Naturally, an organization has some general duties from the start, and those acting on behalf of the organization can establish new obligations for the organization through their acts, for instance by establishing contracts with other agents (persons, organizations, etc.). And such duties of the organization must be distributed among the different positions, specifying the norms that apply to those that occupy such positions (that hold such roles), and usually attributing to them the power to act in the name of the organization, with respect to the fulfilment of such duties. And in this way we have a *dynamic of obligations*, where the obligations flow from the organization to the holders of some roles, and these, through their acts, create new obligations for the organization.

The organization's statute normally distributes the "general duties" of the organization among the different positions, and gives power to the holders of some positions to make a similar distribution of the concrete duties that will be attached to the organization through its normal activity (by the acts of those that act on the organization's behalf).

The allocation of a duty φ of the organization *o* to a role *r* can be described through formulas of the form

$$\mathrm{O}_o\varphi \rightarrow \mathrm{O}_r\varphi$$

i.e., more precisely (since, similarly to what happened before, with respect to the previous "collective obligation", we can interpret an obligation, $\mathrm{O}_{\mathrm{o}}\varphi$, on a "collective agent" *o*, as $\mathrm{OE}_{\mathrm{o:o}}\varphi$)

$$\mathrm{OE}_{o:o}\varphi \rightarrow \mathrm{O}_r\varphi$$

where the attribution of deontic notions (obligations, permissions and prohibitions) to roles is defined by

$$\mathrm{O}_r\varphi \;=_{\mathrm{def}}\; \forall x{\in}\mathrm{X}\;(qual(x{:}r) \rightarrow \mathrm{OE}_{x:r}\varphi)$$

$$\mathrm{P}_r\varphi \;=_{\mathrm{def}}\; \forall x{\in}\mathrm{X}\;(qual(x{:}r) \rightarrow \mathrm{PE}_{x:r}\varphi)$$

$$\mathrm{F}_r\varphi \;=_{\mathrm{def}}\; \forall x{\in}\mathrm{X}\;(qual(x{:}r) \rightarrow \mathrm{FE}_{x:r}\varphi)$$

where *qual*(*x*:*r*) is true if and only if the agent *x* holds the role *r* – "agent *x* is *qualified* to play the role *r*". In general, given a role *r(...)*, we have that *qual*(*x*,*r(...)*) is a predicate that can be described as *is*-*r*(*x*,...). For instance,

qual(*a*,president_of(*o*)) = *is*-president_of(*a*,*o*)

which means that *a* is qualified to play the role of president_of(*o*) iff *is*-president_of(*a*,*o*) is true, that is, iff *a* is the president of organization *o*. (See [9], [27] for details.)

According to the previous definitions, attach an obligation, permission or prohibition, to a role, with respect to some state of affairs φ, corresponds to state that all[51] the holders of the role are obliged, permitted or forbidden, to do φ, acting in the quality of holder of that role. For instance:

- $F_{\text{administrator-of}(o)}\varphi$ informally means that all the administrators of *o* are forbidden to do φ, when acting in the quality of administrators of *o*
- and $OE_{o:o}\varphi \rightarrow O_{\text{president-of}(o)}\varphi$ informally means that the president of the organization *o*, inherits, from *o*, the obligation of doing φ (acting in that quality of president), thus expressing that such obligation, of the organization, flows for its president.

The previous definitions allow the deontic characterization of roles independently from the agents that hold them at a particular moment. Even in the (more or less frequent) cases where we have a role that can have only one holder (as e.g. president-of(*o*)), it is still useful attach deontic notions to the role (defined as above), instead of directly to the current holder of such role, since this one can change.

However, in such cases, of a role *r* with only one holder, it might be useful to consider abbreviations to abstractly refer to his current holder, like for instance *the-r* (e.g. *the*-president-of(*o*), etc.).

Using that abbreviation, and defining (similarly to what we have done before)

$E_{b:r1\ /_s\ a:r}\ \varphi$ (or simply, assuming *s* implicit, $E_{b:r1\ /\ a:r}\ \varphi$)

(read "according to the institution *s*, *b* has done φ, playing role r1, through an act of *a* within role *r*")

as an abbreviation of

$E_{a:r}\varphi \wedge (E_{a:r}\varphi \Rightarrow_s E_{b:r1})$

then, a possible policy of an organization *o*, like "(according to its statute, recognized by the normative system s) the organization *o* always acts through his president", can be expressed by the schema

$E_{o:o}\ \varphi \rightarrow E_{o:o\ /\ \textit{the}\text{-president-of}(o):\ \text{president-of}(o)}\ \varphi$

In [9] we have defined a formal language and a logic where we have formally characterized most of these things (including the formal description of roles). In [27] we have extended such work, including the possibility of defining some relations between roles (namely, implication and incompatibility between roles, and sub-roles), and we have defined a more informal language for the specification of organizations, and interactions between agents (individual or collective), through contracts. We are not going to enter in details here about that.

[51] Note that this does not mean that we cannot define e.g. an obligation that does not apply to a role *r* – that is, to all the holders of *r* – but only to a particular holder *a* of the role r, to do some task φ, within such role. In that case we should write explicitly $OE_{a:r}\varphi$.

3.5 A Little More on Roles

As we have already referred, there exist special relationships that are created between a collective agent (an organization) and other agents, to which are associated norms that describe the desired (ideal) behaviour of the agents engaged in such relationships and the consequences of the acts made by them. To such relationships correspond *roles* that agents can play.

Roles are not agents (although it is natural to informally identify a role with its holder, when this is unique), neither roles are sets of agents, although associated to a role we have the set of agents that are qualified to play such role. Not only the same set of agents may correspond to the set of holders of two distinct roles (e.g. *a* may be the president of two distinct organizations, and in order to know the effects of his acts we must know the role that *a* was playing), as the set of holders of some role may change through time.

Within the context of organized collective agency, roles are used as a high-level mechanism for structuring the desired behaviours, by the association to roles of deontic notions (that describe the obligations and permissions of the agents that can play such roles), powers, etc. But roles should not be confused with their deontic characterization, i.e. roles should not be reduced to mere sets of obligations, permissions or other normative concepts (that apply to the holders of such roles). In particular, the deontic characterization of some role may change with time.

Roles are a very important high-level mechanism to specify how the collective agents are structured and how they behave. But the concept of role, and of acting in a role, is relevant not only within the context of organizations. Roles are fundamental artefacts for understanding and describing agents acting and interacting, in general.

In our opinion, roles may be seen as[52] corresponding to qualities/properties that the agents may have that are relevant for describing agents acting and

[52] For a discussion on the ontological nature of roles see e.g. [25]. According to their authors, in order to some property may be seen as a role, it must satisfy some key features, like anti-rigidity (among others). Using their words: “In general, playing a role is *not a necessity*. Being a Prime Minister is not an essential property of people: for everybody that is a Prime Minister, it would be perfectly possible for her or him *not* to be a Prime Minister (anti-rigidity)...”. Although most of the relevant roles might have this property, we do not think it is essential. For instance, although "to be a King" is not properly a choice of the actual person that is King, we can say that, when performing certain kind of acts, such person exercises the role of King. As another example, *a* may be the father of *b* (and once he becomes father of *b*, he will be father of *b* forever) and at the same time *a* may be a professor at the school of *b* (although not his teacher); and, according to the regulation of the school, *a* might not have right to access to the classifications of *b* as a professor of the school, although he might have access to them in the role of his father. A more detailed discussion of this issue is outside the scope of this paper.

interacting. Such qualities may express properties of an agent, independently of others (like be professor, owner of a building, etc.), or relationships with other agents (like be "president-of", "associated-of", "professor-of", etc.).

Naturally, in practice, we do not associate roles to all properties that agents might have. We only associate a role to a property if the fact that someone has that property may be relevant for some of his acts (e.g. because these are permitted only for persons having that property): e.g., only the owner of "building xpto", or some representative of him, can sell "building xpto"; and only the ones that have the property of being administrator of company *o* are permitted to perform certain kinds of acts.

3.6 Recognition of an Act as an Act in Some Role

In order that an agent acts, playing some role, he or she must be qualified to play that role. In [9], [27] we have expressed this basic principle of our approach by the schema

$$(*) \qquad E_{x:r}\varphi \rightarrow qual(x{:}r)$$

As particular instances of this principle, we have e.g.

$$E_{a\text{:president-of}(o)}\varphi \rightarrow is\text{-president\_of}(a,o)$$

that means that in order to *a* do φ, playing the role of president of the organization *o*, agent *a* must be the president of the organization *o*.

But, as we have mentioned, an agent can be the holder of different roles within the same organization or in different organizations (possibly being subject of potentially conflicting obligations), and can *act playing different roles*. And in order to know the (legal, institutional, ...) effects of his acts we must know in which role they were done. Therefore, *it is fundamental to know which acts count as acts done in a particular role*.

On the other hand, *in reality what we have is agents directly acting* (which we are expressing through formulas of the form: $D_a\varphi$). Thus, *the precise question seems to be*: in what conditions some acts will be recognized as acts in some role, by the environment, the organization, the society, the normative system, etc? And the answer seems to be that there are conventions that make that an act is interpreted as an act playing some role. The recognition of a direct act as an act in some role r follows from some information related with such act (e.g. from some conventional signs exhibited by the agent), according to some common practice or according to some policy of the relevant organization (sometimes just implicit). Here we will give just some examples.

We may have a (possibly implicit) rule that states that someone that is a notary, when performing certain kinds of acts, like signing legal documents, always acts in the quality of notary (for the normative system s). A particular instance would be

$D_a\, \varphi \;\wedge\; \mathit{is}\text{-notary}(a) \Rightarrow_s \; E_{a:\text{notary}}\, \varphi$

Analogously, we may have that any (relevant) state of affairs φ, brought about by the president of an organization o, inside its building, counts as if it was done by him, acting on the quality of president:

$D_a\, \varphi \wedge \mathit{is}\text{-president-of}(a,o) \wedge \mathit{is}\text{-in-the-building-of}(a,o) \Rightarrow_s$

$E_{a}\text{:president-of}(o)\; \varphi$

As other examples, if someone in the street, dressed as a policeman, ask by our identification card, we assume that he is a police officer, acting in that role. And probably any act of the person that is the President of Portugal will be considered as an act playing that role, unless it is an act made in his private home.

In some cases we may even have that an act of an agent *a* will be considered as an act, in any role he can play[53], which we can try to capture by

$D_a\, \varphi \wedge \; \mathit{qual}(a{:}r) \Rightarrow_s \; E_{a:r}\, \varphi$

or that such is the case, but only when is also permitted to do φ, when acting within such role:

$D_a\, \varphi \wedge \; \mathit{qual}(a{:}r) \wedge P_r\varphi \Rightarrow_s \; E_{a:r}\, \varphi$

4. Conclusions

We have combined deontic operators, (various) action operators, of what we called the “brings it about” type, and the counts-as operator. We have illustrated the expressive power that we can get trough these combinations, and have showed how they can be used to represent some concepts that are essential for the understanding of (organized) collective agency, and agents acting and interacting in general.

The proposed level of abstraction provided by these operators seems to be appropriate when we want to model and characterize humans and organizations acting and interacting, at an abstract level, where we do not know yet, or we do not care, about the exact type of actions that can be executed, and where we want to concentrate on the characterization of how obligations flow from the organization to the holders of some positions, and how the acts of these count as acts of the organization.

However, the proposed level of abstraction, providing no resources for representing and reasoning about the temporal dimension, the effects of state change and specific actions, also has several limitations.

For instance, the logical characterization (*) of the principle "for an agent to act, playing a certain role, he or she must be qualified to play that role", given in the previous section, works well, in general, because in most cases

[53] Or an attempt to act, similarly to what it is considered in e.g. [1].

the qualification of the agent to play that role does not change in consequence of his mentioned act. When this is not the case, we may have problems with such representation. Suppose e.g. that an agent *a*, playing the role of owner of building xpto, sells the building. Then, by (*) we have that

$$E_{a\text{:owner-of(xpto)}}\text{sells(xpto)} \rightarrow \textit{is}\text{-owner\_of}(a\text{,xpto})$$

but this makes no sense! After *a* having sold the building xpto, *a* is no longer the owner of that building. In these cases it seems that the real relevant moment to evaluate the qualification is immediately before the act! Similar problems may occur, in some examples, when we try to characterize the conditions under which a direct act is interpreted as an act playing some role.

In these cases, we would like also to be able to express the state change associated to the agents' acting, similarly to what is provided by dynamic logic. This suggests a combination of these action operators with a kind of dynamic logic approach, a topic that was addressed in [7], within the stit semantic.

However, more work is needed, both in the formal characterization of these combinations of operators, as in their illustration, within practical relevant examples.

Power and leadership are related concepts. Herein we have discussed some power notions, and how to represent them, but we omitted discussion of leadership. Leadership expresses a relation of influence, a kind of "influencing power", and typically it is supposed to be associated to (the holders of) certain positions, within organizations. What are the characteristics inherent to potential leaders? Leaders must always be innovators, with a strategy behind? Leaders' characteristics depend on the contexts, and differ e.g. from government organizations (political contexts) to other kind of organizations? How fundamental is trust in the relationship with the leaders? These and other related questions were not addressed in this paper, but are relevant to our general topic and deserve further research.

Acknowledgements. A preliminary version of this paper was presented at the "Arrabida Workshop on Power, Leadership and Complexity" that took place at July 2-4, 2012. The author would like to thank João Caraça and Helder Coelho for their invitation to give this talk, and the participants of the workshop by their comments and discussions. A special thanks in this regard is due to Cristiano Castelfranchi.

Parts of this work have been also presented previously, as invited talks, at the "Eighth International Workshop on Deontic Logic in Computer Science" (DEON-06), held in Utrecht, The Netherlands, and at the "International Workshop on Philosophy and Ethics of Social Reality 2007" (SOCREAL 2007), held in Hokkaido University, Sapporo, Japan. The author would like to thank, also, respectively, J.-J. Meyer and Lou Goble, and Tomoyuki Yamada for the invitation to give those talks, and the participants of these workshops (and in particular Mark A. Brown) by their criticisms and discussions on some of the topics of this work.

Finally, the author wants to gratefully acknowledge the support from FCT, project PEst-OE/MAT/UI0219/2011.

References

1. Abadi, M., Burrows, M, Lampson, B., Plotkin, G.: A Calculus for Access Control in Distributed Systems. ACM Transactions on Programming Languages and Systems, Vol. 15, No. 4, 706-734 (1993)
2. Belnap, N.: Backwards and Forwards in the Modal Logic of Agency. Philosophy and Phenomenological Research, 51, 777-807 (1991)
3. Belnap, N., Perloff, M.: Seeing To It That: A Canonical Form for Agentives. Theoria, 54, 175-199 (1989)
4. Brown, M.A.: Agents with Changing and Conflicting Commitments: A Preliminary Study. In: Prakken, H., MacNamara, P.F. (eds.) Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science, pp. 109-125. IOS Press (1999)
5. Brown, M.A.: Obligation, Contracts, and Negotiation: Outlining an Approach. Journal of Applied Logic, 3, 371-395 (2005)
6. Brown, M.A.: Acting with an End in Sight. In: Gouble, L., Meyer, J.-J.Ch. (eds.) Deontic Logic and Artificial Normative Systems. LNAI, vol. 4048, pp. 69-83. Springer, Heidelberg (2006)
7. Carmo, J.: Collective Agency, Direct Action and Dynamic Operators. Logic Journal of IGPL, vol 18, nº 1, 66-98 (2010)
8. Carmo, J., Jones, A.J.I.: Deontic Logic and Contrary-to-Duties. In: Gabbay, D.M., Guenthner, F. (eds.) Handbook of Philosophical Logic, 2nd edition, vol. 8, pp. 265-343. Kluwer, Dordrecht (2002)
9. Carmo, J., Pacheco, O.: Deontic and Action Logics for Organized Collective Agency, Modeled through Institutionalized Agents and Roles. Fundamenta Informaticae (Special Issue on Deontic Logic in Computer Science), 48 (2, 3), 129-163 (2001)
10. Castelfranchi, C.: The Micro-Macro Constitution of Power. In: Tuomela, R., Preyer, G., Peter, G. (eds.) An International Journal of Interdisciplinary Research, Understanding the Social II - Philosophy of Sociality, Double Vol. 18-19, pp. 208-265 (2003)
11. Chellas, B.J.: The Logical Form of Imperatives. Dissertation, Stanford University (1969)
12. Chellas, B.J.: Modal Logic – An Introduction. Cambridge University Press (1980).
13. Chellas, B.J.: Time and Modality in the Logic of Agency. Studia Logica, 51, 485-517 (1992)
14. Elgesem, D.: Action Theory and Modal Logic. Dissertation, University of Oslo (1993)
15. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. In: Gabbay, D.M., Guenthner, F. (eds.) Handbook of Philosophical Logic, 2nd edition, vol. 4, pp. 99-217. Kluwer, Dordrecht (2002)
16. Herrestad, H.: Formal Theory of Rights. PhD thesis, Dept. of Philosophy, University of Oslo (1996)
17. Hilpinen, R. (ed.) Deontic Logic: Introductory and Systematic Readings. Reidel, Dordrecht (1971)
18. Hilpinen, R.: On the Semantics of Personal Directives. Ajatus, 35, 140-157 (1973)
19. Hilpinen, R.: On Action and Agency. In: Ejerhed, E., Lindström, S. (eds.) Logic, Action and Cognition – Essays in Philosophical Logic, vol 2 of Trends in Logic, Studia Logic Library, pp. 3-27. Kluwer (1997)

20. Jones, A.J.I., Sergot, M.J.: On the Characterization of Law and Computer Systems: The Normative System Perspective. In: Meyer, J.-J.Ch., Wieringa, R.J. (eds.) Deontic Logic in Computer Science: Normative System Specification, pp. 275-307. Wiley, England (1993)
21. Jones, A.J.I., Sergot, M.J.: A Formal Characterization of Institutionalized Power. Journal of the IGPL, 4 (3), 429–445 (1996)
22. Kanger, S.: Law and Logic. Theoria, 38 (1972)
23. Krogh, C.: Normative Structures in Natural and Artificial Systems. Tano.Aschehoug: Institutt for Rettsinformatikk, Complex 5/97 (1997)
24. Lindahl, L.: Position and Change - A Study in Law and Logic. D. Reidel: Synthese Library 112 (1977)
25. Masolo, C., View, L., Bottazi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social Roles and their Descriptions. In: Dubois, D., Welty, C., Williams, M.-A. (eds.) Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning, pp. 267-277. AAAI Press. (2004)
26. Meyer, J.-J.Ch., Wieringa, R.J.: Deontic Logic: A Concise Overview. In: Meyer, J.-J.Ch., Wieringa, R.J. (eds.) Deontic Logic in Computer Science: Normative System Specification, pp. 3-16. Wiley, England (1993)
27. Pacheco, O., Carmo, J.: A Role Based Model for the Normative Specification of Organized Collective Agency and Agents Interaction. Journal of Autonomous Agents and Multi-Agent Systems, 6, 145-184 (2003)
28. Pörn, I.: The Logic of Power. Blackwell, Oxford (1970)
29. Pörn, I.: Action Theory and Social Science: Some Formal Models. D. Reidel: Synthese Library120 (1977)
30. Royakkers, L.: Extending Deontic Logic for the Formalisation of Legal Rules. Kluwer (1998)
31. Royakkers, L., Dignum, F.: Collective Obligation and Commitment. Paper presented at Il Diritto nella Societa dell'Informazione conference, Firenze (1998)
32. Santos, F., Carmo, J.: Indirect Action, Influence and Responsibility. In: Brown, M., Carmo, J. (eds.) Deontic Logic, Agency and Normative Systems. Workshops in Computing Series, pp. 194-215. Springer: (1996)
33. Santos, F., Jones, A.J.I., Carmo, J.: Action Concepts for Describing Organised Interaction. In: Sprague, R.A.Jr. (ed.) Proc. of the Thirtieth Annual Hawaii International Conference on System Sciences, pp. 373-382. IEEE Computer Society Press, vol V (1997)
34. Sergot, M.J.: Normative Positions. In: MacNamara, P., Prakken, H. (Eeds.) Norms, Logics and Information System: New Studies in Deontic Logic and Computer Science, pp. 289-308. IOS Press, Amsterdam (1999).
35. Sergot, M.J., Richards, F.: On the Representation of Action and Agency in the Theory of Normative Positions. Fundamenta Informaticae, 45, 1-21 (2001)
36. Tuomela, R.: The importance of us. Stanford University Press: Stanford series in Philosophy (1995)