HOKKAIDO UNIVERSITY

Title	Logics for modeling (organized) collective agency
Author(s)	Carmo, José
Description	SOCREAL 2007: International Workshop on Philosophy and Ethics of Social Reality. Sapporo, Japan, 2007-03-09/10. Session 1: Collectivity
Citation	SOCREAL 2007: Proceedings of the International Workshop on Philosophy and Ethics of Social Reality, Sapporo, Japan, 2007 / editor Tomoyuki Yamada, 1(1)-21(21)
Issue Date	2007
Doc URL	https://hdl.handle.net/2115/29934
Type	conference paper
File Information	carmo.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

SOCREAL 2007
International Workshop on Philosophy and Ethics of Social Reality 2007
Hokkaido University, Sapporo Japan

Logics for modeling (organized) collective agency

José Carmo
Engineering and Mathematics Department
University of Madeira (Madeira, Portugal)
jcc@uma.pt

My thanks to the GPAE and to *Tomoyuki Yamada*

Part of this talk is based on some joint works with ***Olga Pacheco***

Goal: to use and combine deontic and action modal operators, plus the counts-as operator, in order to characterize some concepts that are essential for the understanding of (organized) collective agency and agent's acting and interacting in general.

We start by making a brief review of these logics, that we want to use, extend and combine.

Notation used:
Capital Latin letters (D, E, F, H, O, P) will be reserved for modal operators; *a*, *b*, ... for agents; φ, ψ, ... for sentences (assertions about the states of affairs, etc.); α, β for actions; *s* for the relevant normative/legal system (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). Propositional (Boolean) connectives: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$.

Part I - Preliminaries: modal logics we want to use and combine

Deontic logic

- The traditional approach sees deontic logic as a branch of modal logic

1. the necessity operator is interpreted as meaning *obligation*, and denoted by O
2. the dual of O (¬O¬) is then denoted by P and interpreted as meaning *permission*
3. and *prohibition* is expressed by an operator F, defined as O¬

Standard deontic logic (SDL, for short)

- The models are the *standard modal models M=(W,R,V),* where

  $W \neq \varnothing$ is the set of worlds (or states)

  R is the accessibility relation (assumed serial): wRv means v is an *ideal version* of w

  V assigns to each atomic sentence p the set of worlds where p is true

  $M \models_w O\varphi$ iff $\forall_v$(if wRv then $M \models_v \varphi$)

- We get O as a *normal KD modality* (using Chellas classification), that is:

  (PC) All instances of tautologies

  (K) $O(\varphi \rightarrow \psi) \rightarrow (O\varphi \rightarrow O\psi)$

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

  O-necessitation (rule): if $\vdash \varphi$, then $\vdash O\varphi$

  plus Modus Ponens (MP)

SDL gives raise 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 $\vdash O\varphi \rightarrow O\psi$

And many other proposals have been made to try to solve the paradoxes:

- ✓ combine two accessibility relations (an *ideal* and a *sub-ideal*)
- ✓ consider ideal and sub-ideal worlds + ideal and sub-ideal transitions between worlds
- ✓ consider e.g. variants of the *minimal models* of Chellas of the form: $M=(W,f_O,V)$, where $f_O:2^W \rightarrow 2^W$ ($f_O(Z)$ denotes the set of worlds where proposition Z is obligatory) and

$$M \models_w O\varphi \text{ iff } w \in f_O(\|\varphi\|) \quad \text{where} \quad \|\varphi\| = \{v : M \models_v \varphi\}$$

- ✓ consider dyadic obligation operators and a temporal dimension, or a preference (ideality) ordering of the worlds, or contexts

In this talk we do not have time to enter in details on this issue
(also the deontic component, in isolation, is not our main focus here)

There have been also proposals for defining deontic logic on the top of **dynamic logic** - briefly:

- Dynamic logic appeared in Computer Science related with the correctness proofs of programs.
- In dynamic logic, associated to each action α there exists a modal operator $[\alpha]$, and expressions of the form $[\alpha]\varphi$ are read "(if α is executed, then) after action α, φ is the case"
- Models: $M = \langle W, \{R^\alpha$: for each action $\alpha\}, V\rangle$

$$w\ R^\alpha\ v \quad \text{iff} \quad \text{we can obtain v executing } \alpha \text{ on w}$$

$$M \models_w [\alpha]\varphi \quad \text{iff} \quad \forall_v (\text{if } w\ R^\alpha\ v \text{ then } M \models_v \varphi)$$

The relations R^α associated to some forms of composition of actions (like sequencing, choice and iterations) are then constrained in order to get the intuitive desired meaning. Particular complications occur when we consider actions (like some programs) that may not terminate, or when we consider an action that corresponds to the "negation of an action α "

- J.-J. Meyer proposed to define deontic logic on the top of dynamic logic, as follows:

$$F(\alpha) \leftrightarrow [\alpha]V \quad , \quad P(\alpha) \leftrightarrow \neg F\alpha\ (\leftrightarrow \langle\alpha\rangle\neg V) \quad \text{and} \quad O(\alpha) \leftrightarrow F(-\alpha)\ (\leftrightarrow [-\alpha]V)$$

where V denotes a special violation atom, $\langle\alpha\rangle = \neg[\alpha]\neg$ and $-\alpha$ (the "negation" of α) expresses the non-performance the action α. *(The deontic operators are applied to action terms)*

Preliminaries: Action logics
The "brings it about" action/agency logic

- Kanger, Pörn and Lindahl have combined deontic and action logics as basic building blocks to describe social interaction and complex normative concepts. Their logics have sufficient expressive power to be able to articulate several distinctions at an appropriate abstract level, mainly in virtue of the modal logic of action they employ.
- They introduce an action operator (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:

  "the agent *a* brings it about that φ" ("agent a has brought 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"

- For the characterization of the E_a operator we find different approaches:

1. Kanger and Pörn define E_a as Boolean combinations of two normal modalities.

   Pörn's proposal can be seen as follows: consider standard models with two accessibility relations, associated to each agent *a*, $R1_a$ and $R2_a$:

   $(w,v) \in R1_a$ iff everything which *a* brings about in *w* is the case in *v*

   $(w,v) \in R2_a$ iff not everything that *a* brings about in *w* is the case in *v*

   $M \models_w E_a\varphi$ iff $\forall_v$(if $w\,R1_a\,v$ then $M \models_v \varphi$) and $\exists_v(w\,R2_a\,v$ and $M \not\models_v \varphi)$

   Idea: $E_a\varphi$ iff it is necessary for something *a* does that φ, but for *a*'s action it might have been the case that ¬φ

2. Dag Elgesem and others consider E_a as primitive, and define its semantics using variants of the minimal models (e.g. Santos & Carmo consider, for each agent *a*, a function f_a such that $f_a(Z)$ denotes the set of worlds where agent *a* brings it about *Z*)

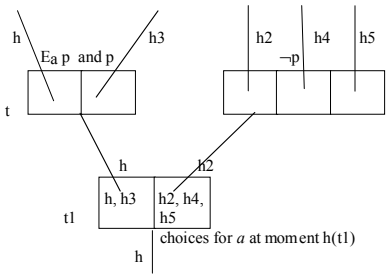
3. A more elaborated semantics can be find within the "stit theory" of Nuel Belnap and Michael Perloff:

They use $STIT_a\,\varphi$ instead of $E_a\varphi$ and for them "agent *a* sees to it that φ" if "the present fact that φ is guaranteed by a prior choice of *a*". Their models use histories, time, moments and agent's choices:

$M \models_{h,\,t} E_a\,\varphi$ iff there is a time (*a choice point)* $t1 < t$ such that

(1) $M \models_{h1,\,t} \varphi$ for every history h1 belonging to the same choice set of h at time t1

and (2) $M \not\models_{h2,\,t} \varphi$ for some history h2 with the same past and present that h at time t1



Drawback: According to Belnap's initial proposal, the schema $\neg E_a E_b\,p$ (for $a \neq b$) is valid,

which excludes (at least some) notions of direct control

- Axiomatic presentation

With few exceptions, all the approaches sees E_a as a non-normal modality, satisfying

(Tautologies and Modus Ponens +)

(T) $E_a\varphi \rightarrow \varphi$ (E_a is a "success" operator)

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

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

\+ (No) $\neg E_a\,\top$ (where ⊤ denotes a tautology)

Schema (No) is used 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 is the result of actions by agent *a*. No agent can meaningfully bring about what is logically true or, more generally, what was unavoidable.

- The "brings it about" approach to the logic of action offers an expressive power rather different from others action logics (that is particularly useful in some contexts):

✓ one 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 $E_aE_b\varphi$ (*a* made *b* do), $E_a\neg E_b\varphi$ (*a* made *b* avoid)

✓ Moreover we can combine the action operators with the deontic operators and use that to express various kinds of normative relations between agents.

Suppose e.g. that *rmr(x,y)* denotes "*x* reads medical record of *y*" , *d* denotes a doctor and *p* a patient. Suppose also that according to some Hospital regulation "Patients do not have the right to access to his or her own record". What does this exactly mean ?

$PE_d rmr(d,p) \wedge \neg PE_p rmr(p,p)$ or $PE_d rmr(d,p) \wedge \neg PE_p rmr(p,p) \wedge OE_d\neg E_p rmr(p,p)$ or ...

Stig Kanger and Lars Lindahl have used these kind of combinations to express legal concepts and relations like rights, duties, etc, and Ingmar Pörn used them to the study of "control and influence" relations in social interactions

- Comparing with the dynamic logic approach
- the dynamic operator $[\alpha]$ is a kind of conditional operator, whereas E_a is a "success" operator;
- $[\alpha]$ is centered on the actions, whereas E_a is centered on the results (abstracting from the particular actions performed);
- $[\alpha]\varphi$ is evaluated before the hypothetical execution of the referred action, whereas $E_a\varphi$ is evaluated after the execution of the relevant action(s);
- Dynamic logic is particularly suited to describe state changes and the effects of actions, whereas E_a is static
- Central to the "brings it about" concept is the notion of agency and causation / responsibility.

- Some extensions and refinements of these action logics:

✓ Santos, Jones & Carmo proposed a non-necessarily successful action operator H_a, with the following informal meaning:

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

(Although we may bring about something without intention, the concept of attempting clearly presupposes intention.)

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

We may combine both operators on actions.

For instance, we may say that an agent *a* has an effective (or total) control with respect to the state of affairs φ, if $(H_a\varphi \rightarrow E_a\varphi)$ is the case.

- Some other extensions and refinements of these action logics:

✓ Sometimes it is also useful to distinguish between the direct and immediate effect's of agent *a*'s actions, and those indirect effects that follow from such direct effects, either sometime later by some causal connection, or by institutional connection (or by other reasons), effects that can also be attributed to the responsibility of *a*'s actions (and described through the E_a operator),

Herein we will use $D_a\varphi$ to state that

"agent *a* has directly (just) brought it about that φ is the case"

and we will continue to use the operator E_a to express states of affairs that are generically brought about by agent *a*, in any generic broad sense.

Preliminaries: Counts-as

- Andre Jones & Marek Sergot (1996):

  "within institutions, organizations, or other normative systems, there are rules that state that some acts, or some state of affairs *count as*, or *are to be classified* as acts, or state of affairs of a different kind (rules that may differ from system to system)"
- To express such *count as* ("*meaning*") relations they propose the use of a conditional modal operator, denoted by $\Rightarrow_s$ (where the index s refers the relevant system of analysis)
- Although they do not restrict the use of $\Rightarrow_s$ connected to act-descriptions, that is their main goal:

  $$E_a\varphi \Rightarrow_s E_b\psi$$

  "*a*'s act of bringing about φ counts, within 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 of, or as an agent of *b*"

- Jones & Sergot also introduce a normal modal operator, D_s, where $D_s\varphi$ may be read as follows: "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 that "according to the rules operating/accepted in institution *s*, φ is the case" or "it is recognized/accepted by institution *s* that φ is the case"
- For $\Rightarrow_s$ they proposed a conditional logic of type CE, containing the rules of replacement of logical equivalents both in the antecedent and in the consequent, plus the following schemas (but not their converses):

  $\vdash ((\varphi \Rightarrow_s \psi 1) \wedge (\varphi \Rightarrow_s \psi 2)) \rightarrow (\varphi \Rightarrow_s (\psi 1 \wedge \psi 2))$

  $\vdash ((\varphi 1 \Rightarrow_s \psi) \wedge (\varphi 2 \Rightarrow_s \psi)) \rightarrow ((\varphi 1 \vee \varphi 2) \Rightarrow_s \psi)$
- For the operator R_s they propose it as a normal modality of type KD, rejecting both

  the T-schema $R_s\varphi \rightarrow \varphi$ and its converse $\varphi \rightarrow R_s\varphi$
- And as bridging principles, they propose

1. $(\varphi \Rightarrow_s \psi) \rightarrow R_s(\varphi \rightarrow \psi)$
2. they also adopt $(\varphi \Rightarrow_s \psi) \rightarrow (\varphi \rightarrow R_s\varphi)$

   as a means of securing the stronger detachment principle $(\varphi \Rightarrow_s \psi) \rightarrow (\varphi \rightarrow R_s\psi)$

Deontic logic again:

It is natural to try to use deontic operators to direct individual personal agency (by forbidding some actions, making others obligatory, etc.), and we could conceive explicit personal deontic operators, like:

$O_a\varphi$: *a* is under an obligation of doing φ

$P_a\varphi$: *a* is permitted to do φ

$F_a\varphi$: *a* is forbidden to do φ

- A natural question is if such personal deontic operators need to be primitive, or can be defined in terms of other operators. As we have seen, we can combine and iterate impersonal deontic operators and action operators, and it seems natural to define:

  $O_a\varphi = O\,E_a\varphi$

  $P_a\varphi = P\,E_a\varphi$

  $F_a\varphi = F\,E_a\varphi$

- However, we should be aware that these definitions have some consequences:

  First we loose the interdefinability between the personal deontic operators:

  We still get $P_a\varphi \leftrightarrow \neg F_a\varphi$ as well as $O_a\varphi \rightarrow P_a\varphi$ (if O satisfies the D-schema), *but*

  $F_a\varphi \leftrightarrow O_a\neg\varphi$ is no longer valid (since $O_a\neg\varphi = OE_a\neg\varphi$ is stronger that $F_a\varphi = O\neg E_a\varphi$)

  We do not think that this is a problem !

- But we will have problems, if O verifies the

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

- Main problem - the problem of transmission of obligations:

✓ Since (by the (T) schema) $\vdash E_aE_bB \rightarrow E_bB$

by the (RM)-rule: $\vdash O_aE_b\varphi \rightarrow O_b\varphi$

(which is clearly unacceptable)

✓ As well as, since $\vdash E_a\varphi \rightarrow \neg E_b\neg\varphi$

by the (RM)-rule: $\vdash O_a\varphi \rightarrow F_b\neg\varphi$

making impossible to express conflicts of obligations between different agents.

What the deontic status of *a* has to do with the deontic status of *b* ?

- However, as we have already referred, we can define the O operator in such a way that it does *not* verify the (RM)-rule (possibly satisfying some weaker versions of it)

- But how to define and characterize deontic operators when we want to use them to direct the agency of groups of agents, organizations, etc. ?

Part II - Organized Collective Agency
Agents acting

- Agents may be subject of obligations and in order to fulfill 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 making contracts).
- And, as we have already seen, we may combine deontic and action operators (for instance of the brings it about type) to describe the obligations that apply to each agent, and the effects of his actions on the relevant state of affairs.
- But if we want to describe social interaction we cannot avoid joint actions and collective agency.

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 we can extend the brings it about operator to a set X of agents in order to capture that (as was done by Lindahl):

  $E_X\varphi$: the set of agents described in X jointly cooperate to bring about that φ is the case
- In this way we can express some notions of *collective agency*, and define logics where formulas of the form $E_{\{a,b\}}\varphi \wedge \neg E_a\varphi \wedge \neg E_b\varphi$ (with $a \neq b$) can be consistent
  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

  $E_X\varphi \rightarrow E_Y\varphi$, for $X \subseteq Y$
- Using this operator, we can describe the establishment of contracts. For instance, *a* and *b* may establish a contract between them, by which *a* becomes under the obligation of doing φ and *b* becomes under the obligation of doing ψ: $E_{\{a,b\}}(O_a\varphi \wedge O_b\psi)$

Collective agency

- Suppose now that a set X of agents wants to act collectively in a more or less permanent basis
- First hypothesis: whenever the group X wants to act, all the members of X meet and jointly act
- In such case we could use the previous action operators to express such situations.
- But, if we assume that the group X wants to act collectively in a more or less permanent basis, probably it will interact with other agents and groups, making contracts, etc., and in this way creating obligations on the proper group X of agents.
- The question now is how we can characterize such kind of *collective obligation*: O_X
- Since we are assuming that the group always act through a joint act of all its members, it is natural to assume that the 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.

  We defend that only 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, such collective obligation must be defined in terms of obligations on the agents, and, at a first sight, in terms of individual obligations of the members of X.

  Two options seem then apparently natural

1) A "general (or universal) obligation", where an obligation on the group corresponds to an obligation on each of its member $O_X\varphi = \forall_{ag\in X}\, O_{ag}\varphi$
2) A "weak general obligation", where an obligation on the group corresponds to an obligation on some of its members $O_X\varphi = \exists_{ag\in X}\, O_{ag}\varphi$

- Option 2) does not serve : Not only it validates $O_X\varphi \rightarrow O_Y\varphi$, for $X\subseteq Y$, as it is of little interest in practice. If the obligation is not fulfilled, who is the responsible ?
- In the case under analysis, where group X always act through the joint agency of all its members, option 1) seems a good option, *but* we must be careful in stating the obligation that applies to each of its members. We do *not* want to say that $\forall_{ag\in X}\, OE_{ag}\varphi$ (suppose e.g. that X is a football group/team and φ is "to mark (at least) five goals on today's game")
- The most natural interpretation of $OE_X\varphi$ would be, in this case

  $\forall_{ag\in X}\, OE_{ag}E_X\varphi$ or, using the attempt operator, $\forall_{ag\in X}\, OH_{ag}E_X\varphi$

Collective agency – Acting in the name of, counts-as and direct acts

- Continue to suppose that a group X wants to act collectively in a more or less permanent basis
- However, the previous case where the group always act through a joint act of all its members is not the most usual case
- Usually, the group will organize its activities in some stable way, allowing that some acts may be performed in the name of the group by some of its 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) its member *a* may bring about certain type of states of affairs φ in the name of the group;

ii) or that any of its members may bring about certain type of states of affairs φ in the name of the group (suppose that X is a group of Mafia killers and they have a code that "when one kills, all kill")

- If we want to characterize such situation, what it will be required ?

- One hypothesis would be to state:

(*i) $E_X (E_a\varphi \rightarrow E_X\varphi)$ and (*ii) $E_X \forall_{ag\in X} (E_{ag}\varphi \rightarrow E_X\varphi)$

- However, these formalizations have some imprecision's and drawbacks:

Suppose we have: (*ii) , $a \in X$, $E_a\varphi$ and φ=is-dead(*b*).

Should we derive that E_X is-dead(*b*) ?

First, $\forall_{ag\in X} (E_{ag}\varphi \rightarrow E_X\varphi)$ may describe only an internal agreement of the members of X, not necessarily accepted by the "external world" (the "society" *s*).

The normative system may not recognized that $E_X$is-dead(*b*) is the case, if *a* has killed *b* alone. (The situation is not the same as when $E_X OE_a$is-dead(*b*) and *a* kills *b*).

- In such case we should replace the material implication by the counts-as operator

(**i) $E_X (E_a\varphi \Rightarrow_X E_X\varphi)$ and (**ii) $E_X \forall_{ag\in X} (E_{ag}\varphi \Rightarrow_X E_X\varphi)$

Then, when $E_a\varphi$ is the case, although we cannot derive that $E_X\varphi$ is the case, we can derive $R_X E_X\varphi$, where an expression like $R_X\psi$ could be read as follows: "(according to the rules accepted by X – by the members of X) , it is recognized/accepted by X (interpreted as it is recognized/accepted by all the members of X) that ψ is the case".

- Of course, there may exist cases where the legal system gives power to X to allow someone to act on its name (for instance, by signing an appropriate document, a family X may give power to someone to sell the family's house).

  In such cases we could write e.g.

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

  and if we do not want to talk about different legal systems, we may identify truth (of some kind of statements) with its recognition by the relevant "legal system"(and, for practical purposes, we can dispense the "counts-as" operator) and simply write

  (*i) $E_X (E_a\varphi \rightarrow E_X\varphi)$

- However, there are still other reasons by which the formula $E_a\varphi \rightarrow E_X\varphi$ still does not represent exactly the state of affairs that the group X wants to create.

  In fact, not all acts of *a* are made in the name of the group X and it is possible that *a* may bring about φ for himself, and not in the name of the group X (suppose e.g. that φ represents "buy a Fiat car"), in which case we do not want to derive $E_X\varphi$ from $E_a\varphi$

- An agent can do a similar direct 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 may be not permitted to use that car when on holiday;

  and even if he is permitted do drive that car on holiday, if he has a car accident the responsability 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).

- Thus, it becomes necessary to express the quality in which *a* has acted (the role *a* was playing) when he or she brought about φ.

- Using $E_{a:X}\varphi$ (or $E_{a \text{ as } X}\varphi$) to denote that *a* as brought about φ as if it was X that has acted (*a* as brought about φ *in the name of* X, or *in the role of representative of* X), then we can write (where $E_{X:X}\varphi$ may be read "X as brought about φ as itself)

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

- Naturally, we can question how we know that *a* as brought about φ as himself or in the name of the group X. We will discuss such issue later.
- Another question is how we can discriminate in our language the cases where

1) X brought about φ, because some agent has brought about φ in the name of X
2) from the cases where X brought about φ by itself (by a joint act of the group X)

- We can introduce a notation to express like $E_{X:a}\varphi$ $E_{X:_s a}\varphi$

(or more precisely,)

(read "according to the system s, X has brought about φ through *a*'s acting")

as an abbreviation of $E_{a:X}\varphi \wedge (E_{a:X}\varphi \Rightarrow_s E_{X:X}\varphi)$

- Other option is to consider the direct action operator, D, to describe the direct acts made by an agent (or the direct joint acts made by a group X of agents), and discriminate 1) and 2) as follows:

1) $E_{X:X}\varphi \wedge \neg D_X\varphi$ 2) $E_{X:X}\varphi \wedge D_X\varphi$

- 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 a presence of a group X that always act through a joint act of all its 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 now this kind of collective obligation $O_X\varphi$ should *not* be seen as a general obligation of some form like

$O_X\varphi = \forall_{ag \in X} O_{ag}\varphi$

- Although the "weak general obligation" $O_X\varphi = \exists_{ag \in X} O_x\varphi$ 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 the group X must do (when it creates its statute). The group must guarantee that

$O_X\varphi \rightarrow \exists_{ag \in X} O_{ag}\varphi$

And, for this, 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 the obligation refers to (and also giving representative powers to such acts)

Collective agency – collective agents and roles

- Suppose now that, as before, a group X of agents have common interests and wants to act collectively in a more or less permanent basis, but such intended common activity should proceed even if some member of the group is not (in some moment) anymore interested in it, or when it is possible to aggregate later to the group other members.
- In such cases, the natural way for the group X to proceed is to create an entity with a *proper identity* (like it is the case of an organization). The members of the group will be related with such entity by *special relationships*, like member-of (associate-of, or whatsoever), but such entity will persist even when the set of its members will change.
- Of course, such entity that is created by the group X needs to act, and so it is an agent. We have called it an "institutionalized agent; herein we will call it a "collective agent" (even if it may happen that there exists only a unique person related with it).
- In some cases (like when we talk about organizations), the Law will recognize such entity as a "real" agent (an "artificial person"), having juridical personality and legal competence, as any natural person. In other cases, that we do not want to exclude here, such "collective agent" may be more informal, without a legal recognized status.

- Naturally, such "collective agent" needs to act, but *it cannot act directly* ! Thus, someone needs to act on its name. When some group X creates a "collective agent" *o*, a statute is also created where the main normative rules of *o*'s activity are defined, and where, in particular, it is stated the rules by which someone can act in the name of the "collective agent".
- 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 *posts*, or *roles* within the organization (we may call them *structural roles*, 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, in the sense that it does not say which particular person can act in the name of the organization; it attributes such power to the holders of some roles (the titulars of some posts), that have a representative nature.
- Normally, there is an individual post (in the sense that have only one holder), like *president-of*, that can act in the name of the organization, and when the organization it is created not only the statute defines the rules for electing/choosing the holder of such position, but it also defines a (usually provisory) initial titular of such position. The statute also defines if the titular of such post has power to delegate the power of acting in the name of the organization, and in which conditions.

- As we said, to some roles are associated representative powers.

  Suppose that r : REP(o, φ) means that "the role r is a representative role (a role with representative powers) of the collective agent o, with respect to a state of affairs φ (the scope of the representation)".

  The notation r:REP(o, φ) does not mean that the role r can act in the name of o
- We think that only agents can act, and roles are not agents (thus roles do not act)

  What r:REP(o, φ) does mean is that when someone, playing the role r, brings about φ, that can be seen as an act in the name of o, which can be expressed by

  $$\forall_{ag} (E_{ag:r}\varphi \rightarrow E_{o:o}\varphi)$$

  where $E_{ag:r}\varphi$ means "*ag* acting in the role r brings it about that φ" and $E_{o:o}\varphi$ means "o acting in the role of itself brings it about that φ"
- Using the counts-as operator, we can reformulate the previous definition as follows, assuming that such representative status is recognized by the "society" s (the relevant normative system), and write

  $$\forall_{ag} (E_{ag:r}\varphi \Rightarrow_s E_{o:o}\varphi)$$

- Naturally the organization has some general duties from the start, and the agents that can act in the name of an organization may 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 posts, 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 fulfillment 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 in the organization.
- The statute of the organization normally distributes the "general duties" of the organization among the different posts, and give power to the holders of some posts to make a similar distribution of the concrete duties that are attached to the organization through its normal activity (by the acts of those that act in the name of the organization).

- The attribution of a duty of an organization *o* to a particular role *r* of its structure can be described through formulas of the form

  $O_o\varphi \rightarrow O_r\varphi$

  or more precisely (as before, such obligation $O_o\varphi$ on a collective agent *o* should be interpreted as $OE_{o:o}\varphi$)

  $OE_{o:o}\varphi \rightarrow O_r\varphi$

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

  $O_r\varphi =_{def} \forall_{age \in X} (qual(ag:r) \rightarrow OE_{ag:r}\varphi)$

  $P_r\varphi =_{def} \forall_{age \in X} (qual(ag:r) \rightarrow PE_{ag:r}\varphi)$

  $F_r\varphi =_{def} \forall_{age \in X} (qual(ag:r) \rightarrow FE_{ag:r}\varphi)$

  where qual(ag,r) is true if and only if the agent ag holds the role r – agent ag is qualified to play the role r

  In general, given a role r(...) , qual(ag,r(...)) is a predicate that can be described as is-r(ag,...) :

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

- Examples:

  $F_{administrator\text{-}of(o)}\varphi$ informally means that:

  all the administrators of *o* are forbidden of doing φ, when acting in the quality of administrators of *o*

  $OE_{o:o}\varphi \rightarrow O_{president\text{-}of(o)}\varphi$ informally means that:

  the president of *o* inherits, from the organization *o*, the obligation of doing φ

- The previous definitions allow the deontic characterization of roles independently from the agents that hold them in a particular moment.

  Even in the (more or less frequent) cases where we have a role that can have only one holder (like president-of(*o*)), it is still useful to attach deontic notions to the role (defined as above), instead of directly to the actual holder of such role, since this one can change.

- However, in such cases (of a role *r* with only one holder) it still might be useful to consider abbreviations to refer to its actual holder, like (for instance) the_r

- Using such 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 normative system s, agent *b* has brought about φ playing role *r1*, through *a*'s acting playing role *r*") as an abbreviation of

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

  we can express a possible policy of an organization *o*, like "(according to its statute, recognized by the normative system s) the organization *o* always acts through its president", as follows:

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

- In previous works, we have defined a formal language and logic where it is possible to define precisely all these things, including the description of roles and the possibility of defining some relations between roles (incompatibility, sub-roles, etc.), as well as a more informal language for the specification of organizations and interactions between agents (individual and collective) through contracts.

A little more on roles

- Special relationships are created between an organization and other agents, to which are associated norms that describe the ideal expected behavior 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.
- But roles are not agents (although it is natural to informally identify a role with its holder, when this is unique), neither sets of agents, although associated to a role we have the (possible singular) 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 (*a* may be the president of two distinct organizations, and in order to know the effects of his acts we must know the role *a* was playing), as the set of holders of some role may (and usually do) change through time
- Within the context of organized collective agency, roles are used as a high-level mechanism for structuring the desired behaviors, by the association to roles of deontic notions, powers, etc.

  But roles should not be confused with their normative 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 normative characterization of some role may change with time.

- Roles are a very important high-level structural mechanism for abstractly specifying how a collective agent acts and interacts.
- But the concept of role, and of acting in a role, is relevant in many contexts, and not only within the context of the behavior of organizations.
- In our opinion, roles may be seen as corresponding to qualities that the agents might exercise that are relevant when we describe agents acting and interacting with others.

  Such qualities may *either express* properties of an agent, independently of others (like being father, owner of a building, etc.), *or express* relationships with other agents (like being president-of, associated-of, etc.)
- Naturally, in practice, we only associate a role to a property if the fact that someone has that property may be relevant for some of his acts (because, for instance, they are permitted only for persons having that property):

  e.g., only the owner of the building *xpto*, can sell building *xpto*.

Which acts count as acts in a particular role ?

- An agent can be the holder of different roles within the same organization or in different organizations (being the subject of potentially conflicting obligations), and can *act by playing different roles*. And in order to know the (legal, institutional, ...) effects of his acts (and its deontic qualification) we must know in which role they were played.
- Therefore it is fundamental to know which acts *count as* acts 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. ?

  The fact that a direct act is recognized as an act in some role *r* must be deduced from some information related with such act (and the agent involved). For instance:

✓ We may have a (possibly implicit) rule that states that a notary, when performing certain kind of acts φ, like signing legal documents, always act in the quality of notary (for the normative system s):

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

- ✓ 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 brought about by the president, on that quality:

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

- ✓ In some cases, we may have that an act of an agent *a* will be considered as an act in any role he can play *(or an attempt to act, similarly to what is considered within a calculus for access control in DS)*

  $D_a \varphi \wedge \text{qual(a:r)} \Rightarrow_s E_{a:r} \varphi$

  or that such is the case only when it is also permitted to bring about that φ when acting in that role:

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

- In the previous examples, as in most cases, the qualifications of an agent to play some role does not change in consequence of his acts. When this is not the case, it seems that the real relevant moment to evaluate the qualification is immediately before the act, *and* for the recognition of an direct act as an act playing some role what may be relevant is (in some cases) the particular type of action performed and the circumstances when the agent initiated the relevant act.
- In general, there exist some conventional signals that an agent must exhibit in order to guarantee that the act that he will perform will be recognized as an act playing some role. However, these action logics do not provide us the means to talk about such two states (before and after acting).

- One hypothesis is to try to combine these action operators with dynamic operators, and a possibility (among others) is to allow to write expressions of the form $[D_a\varphi]\psi$ informally interpreted as

  $\forall_\alpha [a:\alpha](D_a\varphi \rightarrow \psi)$

- Then we could state, for instance, that

  D_a *doc-own(a,xpto)* $\rightarrow [D_a\varphi]\, R_s\, E_{a:\text{owner-of(xpto)}}\, \varphi$

  where *doc-own(a,xpto)* = "*a* has exhibited the document of ownership of building *xpto*"

  meaning that when an agent *a* shows a document that certificate his ownership of building *xpto*, then any direct (relevant) act that he will made in that moment (like sign a document selling that building) will be recognized (or assumed) as an act playing the role of owner of that building.

- And in this example we could even write that (where $\langle D_a\varphi\rangle = \neg[D_a\varphi]\neg$)

  $\langle D_a\varphi\rangle\, R_s\, E_{a:\text{owner-of(xpto)}}\, \psi \rightarrow D_a$ *doc-own(a,xpto)*

  sentence that would mean that if it is possible for *a* to directly bring it about some state of affairs φ (e.g "sign a document"), that will be recognized as an act in the role of owner of building *xpto*, for bringing about (the same or other possibly different) state of affairs ψ (e.g. "selling xpto"), this means that *a* has directly exhibited the respective document of ownership of *xpto*. This would be a way to express the exact act that is expected as an *authentication* to act as owner of building *xpto*.

Conclusions

- The logical framework here considered, that takes an approach to the treatment of actions which is based on expressions of the form "*a* brings it about that φ is the case ", focus only on the agent concerned and the states of affairs that result from his or her actions. This framework is intended primarily as a formal tool for the characterization of static phenomena.
- The proposed level of abstraction seems to be appropriate when we want to characterize and model human and organizations interaction, 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 and characterize how the obligations flow from an organization to the holders of some posts and how some of the acts of the latter 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 severe limitations, that limits its applicability.
- The combination of these actions logics with dynamic logic operators may be a way to overcome these limitations.

End

and

Arigato gozai masu