Collective action, direct action and dynamic operators

José Carmo

Engineering and Mathematics Department
University of Madeira

jcc@uma.pt

NorMAS 2008

Goal

- Generic goal: to characterize organizations and other normative multiagent systems, seen as entities that are composed by a set of agents and whose behavior is intended to be governed by norms.
- Norms are intended to direct the agency of the different agents, stressing that some acts should be considered as obligatory, permitted or forbidden.
- Thus, it is only natural that for the characterization of such systems we try to use and combine deontic operators, like O, P and F (meaning obligation, permission and prohibition) and action operators.
- The question is: which deontic and action logics we should consider (taking into account the applications we have in mind) ?
- In this talk we want to concentrate on the action logics.

Which action logic ?

- Applications: a first level of specification of organizations where we want to model and characterize 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.
- At such abstract level, generally we do not want to specify the exact means (procedures) that the agents have at their disposal to do the different kind of acts. For many purposes, what is relevant is that an agent a has brought about some state of affairs φ, and he or she is *responsible* for that; or that a is obliged to bring about φ (being possible that a can do it by different means); etc.
- This suggests that we use action logics able to express such *agency* concepts, at this abstract point of view, like the "sees to it"/"brings it about" logics (used by Kanger, Pörn and Lindahl), that introduce a modal operator (generically denoted here by E) that relates an agent (a) with the effects of his action (φ), omitting details about the specific action that was performed.

  The expression $E_a\varphi$ is usually read as "the 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"

“Sees to it”/ “brings it about” action/agency logics and STIT semantics

- Most proposals consider E_a a non-normal modality, satisfying the rule

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

  and including (at least) the schemas

  T: $E_a \varphi \rightarrow \varphi$ (E_a is a “success” operator)

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

  No: $\neg E_a \top$ (No is used to try to capture the concept of *agency* itself)

- With respect to the semantic characterization of this operator, we find different proposals.
- Here we will be concerned with the “stit theory” (stit, for short), of Belnap, Perloff and Horty, where a temporal semantics is provided, trying to capture the intuitions underlying this concept of agency.
- The stit models are based on tree structures, consisting of a nonempty set of moments (Mom), plus a partial ordering $\leq$ satisfying “no downward branching” (at each moment the past is linear and there exist open future branches.)

STIT frames $\mathcal{F}$ = (Mom, ≤, Ag, C) :

- A complete branch in the tree (i.e., a maximal ≤-chain *h*) is called a history, and we say that a history *h* passes through a moment *m* if $m \in h$. H denotes the set of histories and H_m is the set of histories passing through *m*
- Their central element is the notion of agent's choice:

$$C : \text{Mom} \times \text{Ag} \rightarrow \wp(\wp(\text{H})) \quad \text{such that} \quad C(m,a) \text{ is a partition of the set } H_m$$

For each agent *a*, all histories passing through a same moment *m* are partitioned in a set *C(m,a)* of equivalent classes, corresponding to the choices (or actions) that are open/available to agent *a* at the moment *m*.

$C_{m,a}(h)$ (only defined when $m \in h$) denotes the particular action or choice from *C(m,a)* that contains *h*

- In a stit frame the following two conditions are imposed on the choice function *C*:

✓ No choice between undivided histories (if h and h' are undivided at a moment *m*, i.e. there exists $m^* \in h \cap h'$ such that $m < m^*$, then h and h' belong to the same choice set, at *m*, for each agent *a*)

✓ Something happens (informally: no agent can prevent another agent from choosing any of her/his alternatives) The implications of this constraint are discussed in the paper. We can consider or not this constraint.

achievement stit **(*astit*, for short)**

- Within this framework various *stit* operators have been proposed. The most important are:

  the *achievement stit* (*astit*, for short), that tries to capture the results *caused* by the agent's *past choices*, and from which he is *responsible*; and the *deliberative stit* (*dstit*, for short), that tries to capture what is a consequence of the agent's *present choice*.

- Informal idea: an agent *a* sees to it that φ, in the *achievement* sense, which is expressed by [*a astit:* φ], *if* the present fact that φ is guaranteed by a prior choice of *a*, *and* at that choice point it was possible for the agent to make another choice that would not guarantee the current truth of φ (the so called *negative condition*).

- In order to define the *astit* operator, it is considered that the tree can be partitioned horizontally into *instants*, assuming that all branches have a unique temporal order.

  We write *m@h* to denote the unique moment belonging to *h* that is at the same instant as *m*

- A stit model M = (F, *v*) is a stit frame plus a *valuation v* that applies each atomic sentence *p* into the set of pairs *(h,m)* (with $h \in H_m$) at which, intuitively, *p* is true *(which we assume that only depends on m)*

Truth conditions

- Truth is defined with respect to a pair formed by a history h and a moment m (belonging to h):
  $\mathsf{M} \models_{h,m} \varphi$ means that φ is true (in model M) at the moment m, assuming that history h is followed (or according to history h)

- The usual truth-functional connectives are evaluated without changing the index (h,m)
  (e.g. $\mathsf{M} \models_{h,m} p$ iff $(h,m) \in v(p)$; $\mathsf{M} \models_{h,m} \varphi \wedge \psi$ iff $\mathsf{M} \models_{h,m} \varphi$ and $\mathsf{M} \models_{h,m} \psi$; etc.)

- Linear temporal operators can be defined, changing the moment index m through the history h, that is kept fixed (as we will illustrate later)

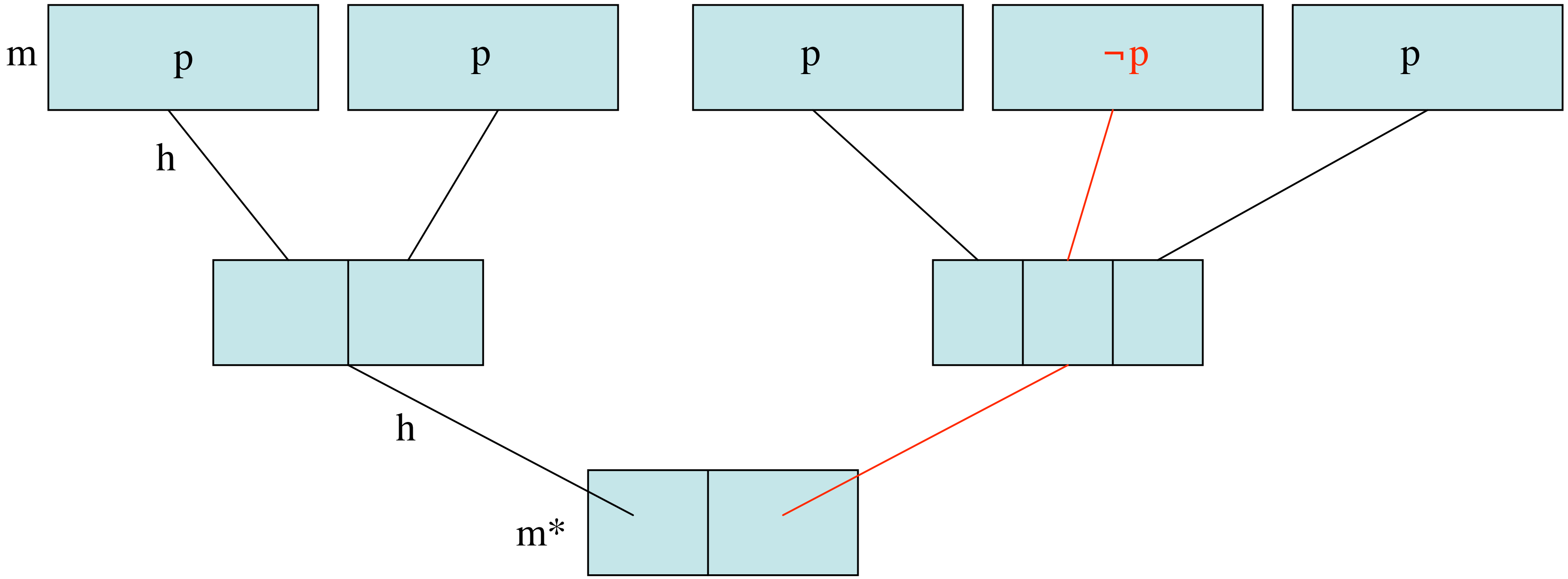
- the following *historical necessity* operator ("settled true at m") is defined changing the history index h and keeping fixed the moment m:
  $\mathsf{M} \models_{h,\, m} \Box\varphi$ iff $\mathsf{M} \models_{h^*,\, m} \varphi$ for *every* history h^* passing through m (i.e. $h^* \in H_m$)

- and $\mathsf{M} \models_{h,\, m} [a\ astit\!: \varphi]$ iff there is a moment $m^* < m$ such that

  1) $\mathsf{M} \models_{h^*,\ m@h^*} \varphi$ for *every* history h^* such that $h^* \in C_{m^*,a}(h)$

(and) 2) $\mathsf{M} \not\models_{h^*,\ m@h^*} \varphi$ for *some* history $h^* \in \mathrm{H}_{m^*}$

- M $\models_{h,m}$ [*a astit:* p] - Example:



the *deliberative stit* (*dstit*, for short) of Belnap and Horty

- M $\models_{h,m}$ [*a dstit:* φ] iff

  1) M $\models_{h^*,\, m}$ φ for *every* history h^* such that $h^* \in C_{m,a}(h)$

  2) M $\not\models_{h^*,\, m}$ φ for *some* history h^* that pass through m (i.e., such that $h^* \in H_m$)

- We can describe [*a dstit:* φ] as follows, using the operator □ and an action operator that expresses only the positive condition 1) above, called *cstit* (from "Chellas stit"):

  M $\models_{h,m}$ [*a cstit:* φ] iff M $\models_{h^*,\, m}$ φ for *every* history h^* such that $h^* \in C_{m,a}(h)$

  and $\models$ [*a dstit*: φ] ↔ [*a cstit*: φ] ∧ ◇¬φ

- We can say that an agent can deliberatively sees to it only sentences about the future.
- The *dstit* sentences concern the agent's present choices / actions, whereas the *astit* sentences refer to the agent's past choices.

  This distinction is important in many respects, e.g. with respect to the meaning of iterations of deontic and action operators:

  What is the informal meaning of $E_a\varphi$ and of $OE_a\varphi$?

  On the other hand, the meaning of the iterations O[*a dstit:* φ] and O[*a astit:* φ] is clear

Joint action

- If we want to describe social interactions, we must also consider joint actions and collective agency.
- Two or more agents can jointly act in order to do some task, like to move a very heavy table, or to make a contract, etc., and the "sees to it" operators can be generalized in order to cover also such situations, as was proposed by Lars Lindahl (although not providing a semantics for these operators). We will extend the *dstit* operator, but the same ideas also apply to the *astit* operator.
- Suppose G denotes a group (a nonempty finite set) of agents.

  Our informal idea regarding this "*joint action*" concept is the following: when we say that the group G jointly sees to it (brings it about) that φ is the case, we want to express that the actions of the agents in G cause φ; the actions of each of such agents were necessary to the production of φ.

  We may say that the agents in G jointly cooperate to bring about φ *(we leave it open if such cooperation was intended or not)*, and are responsible for φ being the case.
- This is different from others stit extensions with groups, like the *strategic* version of stit, where $[G]\varphi$ means "agents in G strategically see to it that φ", that satisfy e.g. $\models [G]\varphi \rightarrow [G^*]\varphi$, for $G \subseteq G^*$

  Here we do *not* want that $[G\ \textit{dstit:}\ \varphi] \rightarrow [G^*\ \textit{dstit:}\ \varphi]$ to be a valid schema, for $G \subset G^*$, since when a group G sees to it that φ, the agents not in G might have nothing to do with the "doing of φ".

dstit joint action

- M $\models_{h,m}$ [*G dstit:* φ] iff

  1) M $\models_{h^*,\, m}$ φ for *every* history h^* such that $h^* \in \bigcap_{b\in G} C_{m,b}(h)$

  2) for *every a* in G, there exists *some* history h^* such that

     h^* pass through m and $h^* \in \bigcap_{b\in G\text{-}\{a\}} C_{m,b}(h)$ and M $\not\models_{h^*,m}$ φ

  (it would be possible to not have φ if any of the agents *a* in G had not acted at *m*, as he or she did (according to *h*), and all the other agents in G have acted as they did: it is only in this case that we can be sure that the actions of each of the agents were really necessary to produce φ)

✓ The definition satisfies what we call *single group* condition: $\models$ [{*a*} *dstit:* φ] ↔ [*a dstit:* φ], as well as

✓ *group anti-monotonicity*: $\models$ [G stit: φ] → ¬[G* stit: φ], whenever G ⊂ G*

- Two simple concrete examples referred by Lars Lindahl (in the "Position and change ..." book):

4 grammes of poison is the minimum to kill person C

At the same time, give to C		A sees to it that C dies		B sees to it that C dies		{A,B} sees to it that C dies	
A	B	Lindahl	Here	Lindahl	Here	Lindahl	Here
2	2	No	No	No	No	Yes	Yes
4	4	Yes	Yes	Yes	Yes	Yes	No

Other forms of "collective agency"

- But we should distinguish between this kind of *collective agency*, that corresponds to a simultaneous direct act of two or more agents, from the agency of entities, that are composed of many agents, like *organizations*.
- An organization is different from the set of agents that are related with it (e.g. these can change, but the organization will keep being the same). We may see an organization as an agent (that we may call of *collective agent*), that have a proper identity, can act and interact with the other agents, making for instance contracts, and can be subject of obligations and be responsible by its violation / non-fulfillment.
- However, these collective agents are different from the other agents by the fundamental fact that they cannot act, directly: someone has to act on its name.
- Relationships are established between an organization and other agents (corresponding to what is usually called of the *posts*, or *roles* within the organization), and there are rules (e.g. in its statute) that:

✓ distribute the duties of the organization among the different posts, specifying the norms that apply to those that occupy such positions (abstractly, that is independently of whom they are at each moment),

✓ settle who has the power to act on its name, specifying that when an agent acts, within some of those *roles*, for bringing about some specific state of affairs, that counts as if it was the organization that has acted.

acting playing some role and direct acting

- Thus, there is 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 in the organization .
- But an agent can be the holder of different roles within the same organization, or in different organizations, and can do a similar act playing different roles, and to know the effects of such act and its deontic classification, we must know in which role it was played.
- For this reason, in previous works, we have *extended* the sees to it operators, expressing the role r that the agent a has played when he, or she, has brought it about that φ, through sentences of the form $E_{a:r}\varphi$.
- But it is not enough to have a way to express the role an agent has played when he has acted. In reality what we have is agents directly acting. Thus, we need also to characterize in what conditions some direct acts will be recognized as acts in some role by the environment, organization, etc.?
- So, it seems important to have means to discriminate the direct and immediate effects of an agent's own actions, from those indirect effects that follow from them, either sometime later, by some causal connection, or (we may suppose immediately) by social convention or institutional connection (by the norms applicable e.g. within the relevant organization), indirect effects that we generically refer through the operator E.

The direct effects of the agent's actions, in the deliberative sense

- Suppose D denotes such direct action operator, and let us see how to characterize such operator, within the stit approach. (Just to simplify, we will consider only the case of one agent).
- Consider first the deliberative stit: we want to express that an agent a has made a choice $C_{m,a}(h)$ (at the present moment m) that has φ (true) as an immediate effect, denoted by

$$\overrightarrow{D}_a \varphi$$

(we do not mean that φ is the case at m (according to h), but that immediately after m, φ is the case.)

- We can define its semantics directly, or follow Mark Brown's approach, and defined it as an iteration of the *dstit* operator and the following linear future operator W (proposed by Mark Brown to capture the notion of *for a while in the immediate future*):

$$M \models_{h,m} W\varphi \text{ iff there } \textit{exists } m^* \in h \text{ such that } m^* > m \text{ and, } \textit{for any } m < m' \leq m^*,\ M \models_{h,m'} \varphi$$

And define: $\overrightarrow{D}_a \varphi$ as [a *dstit:* $W\varphi$]

- Suppose now that we want to express that an agent has *just brought* it about that φ, thus concerning the immediate past actions of the agent (i.e. considering the achievement sense)

The direct effects of the agent's actions, in the achievement sense

- A similar strategy of combining (now) the *astit* operator with a tense operator does not seem obvious, since we need to specify the moment where the relevant action took place, and where the agent could act differently.

  One idea is to implement in the *astit* semantics the following *nearly now* requirement (proposed by Chellas): "it is always *at least the immediate past* that is relevant to whether or not a *STIT* sentence holds at a time in a history: no matter how early a choice is initiated, it continues up to the very moment of agency".

- And one option to try to implement such requirement is to add to the *astit* semantic definition

  $M \models_{h,m} \overleftarrow{D}_a \varphi$ iff there is a moment m* < m such that

  1) $M \models_{h^*,\, m@h^*} \varphi$ for every history h* such that $h^* \in C_{m^*,a}(h)$

  2) $M \not\models_{h^*,\, m@h^*} \varphi$ for some history $h^* \in H_{m^*}$

  a condition like

  3) $\forall_{m^* < m' < m} \forall_{h^*} \forall_{h'} (h^* \in C_{m^*,a}(h) \wedge h' \in H_{m'@h^*} \rightarrow h' \in C_{m'@h^*,a}(h^*))$

  according to which if *a* takes, at *m**, the choice $C_{m^*,a}(h)$, then all choices of *a* between (the choice point) *m** and the time instant of *m* are vacuous choices (*a* does not really act between *m** and *m* instants).

The direct effects of the agent's actions, within discrete time

- It is possible to consider both weaker and stronger conditions than 3), in order to try to capture such *nearly now* requirement.

- Now, we just want to note that if we are concerned with applications where the time is discrete, e.g. isomorphic to the set of integers (or isomorphic to the set of natural numbers), then the things become much simpler.

- In such case the linear operator W is *equivalent* to the "next state" operator X, and using X and the "previous state" operator Y, we can define

✓ $\overrightarrow{D}_a\varphi$ as [*a dstit:* Xφ]

✓ $\overleftarrow{D}_a\varphi$ as Y [*a dstit:* Xφ]

Dynamics

- The sees to it operators are essentially static operators, providing no resources for representing and reasoning about the effects of state change, and (although their virtues) this limits their applicability.
- In particular, we would like to be able to represent and to make hypothetical reasoning about the effects (e.g. institutional) of choosing to do some action, which constitutes an important component within agent decision making. This suggests that we should try to combine these operators with the kind of reasoning that is provided by the dynamic logic.
- Suppose we write $[\vec{D}_a \varphi]\ \psi$ with the following informal meaning: if agent *a* performs (now) an (*any*) action of seeing to it that φ, then, when such action ends, ψ is the case.

  Note that the formula $\vec{D}_a \varphi \rightarrow \vec{D}_a \psi$ does not expresses what we want.

  When we write $[\vec{D}_a \varphi]\ \psi$ we do not intend to state that the agent *a* also sees to it that ψ.

  In particular, we want that our operator will be a normal modal operator.

  Moreover, we want to state that ψ would also obtain, if agent *a* had selected any other action (different from the one expressed by $C_{m,a}$(h)) that would also be an action of seeing to it that φ.

- The desired operator can be defined directly, without considering temporal operators, or as follows:

$$[\overrightarrow{D}_a \varphi]\ \psi \text{ is an abbreviation of } \Box([a\ cstit\!: \overrightarrow{D}_a \varphi] \rightarrow [a\ cstit\!: \mathrm{X}\psi])$$

- Naturally, the two direct action operators become then related by the formula:

$$[\overrightarrow{D}_a \varphi]\ \overleftarrow{D}_a \varphi$$

- And, although we are here mainly interested in applying these kind of dynamic operator [_] to these action sentences, we can generalize and allow it to apply to any sentence (with an agent label), in formulas of the form $[a\!: \varphi]\psi$, with the following informal meaning: "if agent a selects an action that leads to the truth of φ, after such action is performed, ψ is the case" (or, simply, "after a choosing φ, ψ is the case").

  We just have to define: $[a\!: \varphi]\psi$ as $\Box([a\ cstit\!: \varphi] \rightarrow [a\ cstit\!: \mathrm{X}\psi])$

  and we can define $[\overrightarrow{D}_a \varphi]\ \psi$ as an abbreviation of $[a\!: \overrightarrow{D}_a \varphi]\psi$

- And we can define a similar operator without imposing a discrete time (we have just to replace X by W):

$$[a\!: \varphi]\psi \text{ is } \Box([a\ cstit\!: \varphi] \rightarrow [a\ cstit\!: \mathrm{W}\psi])$$

- $[a{:}\ \varphi]\psi$ is a normal with respect to ψ, and, with respect to φ, it verifies the RE-rule, but not the RM-rule, and it satisfies, among others, the following *augmentation principle*: $\models [a{:}\ \varphi]\ \psi \rightarrow [a{:}\ \varphi \wedge \varphi']\ \psi$
- The schema $[a{:}\ \varphi]\psi \rightarrow \psi$ is *not* valid, as it is desirable in order that it can be used for agent decision. For instance, if " a is not liable to punishment" $\wedge$ $[\overrightarrow{D}_a \varphi]$ " a is liable to punishment" is the case, agent *a* should not take the decision of bringing about φ.
- $[a{:}\ \varphi]\psi \rightarrow \varphi$, $[a{:}\ \varphi]\psi \rightarrow \Diamond\varphi$ and $[a{:}\ \varphi]\psi \rightarrow \langle a{:}\ \varphi\rangle\psi$ are also *non* valid schemas
- Note that we must be careful in interpreting $\langle a{:}\ \varphi\rangle\psi$ ($= \neg[a{:}\ \varphi]\neg\psi$), i.e. the dual of $[a{:}\ \varphi]\psi$:

  Although $\langle a{:}\ \varphi\rangle\top$ means that it is possible for agent *a* to choose φ, $\langle a{:}\ \varphi\rangle\psi$ does not mean that it is possible to choose φ in such way that we have a guarantee that after we make that choice, ψ is the case.

  To express that we can consider other kind of dynamic operator, defined as follows:

  $$\langle[a{:}\ \varphi]\rangle\psi \quad \text{is (an abbreviation of)} \quad \Diamond([a\ \mathit{cstit}{:}\ \varphi] \wedge [a\ \mathit{cstit}{:}\ W\psi])$$

  Naturally, $[a{:}\ \varphi]\psi \wedge \langle a{:}\ \varphi\rangle\top$ implies $\langle[a{:}\ \varphi]\rangle\psi$

- As a very simple example of the use of these operators, suppose we want to characterize the conventional signals that an agent must exhibit in order that an act that he will perform will be recognized as an act playing some role *(for instance, the role of owner of a building xpto)*

- Suppose R_s denotes the modal operator proposed by Jones & Sergot, where s denotes an institution, organization, or other normative system, and $R_s\varphi$ reads as follows: "according to the rules/norms operating/accepted in system *s*, φ is the case" or "it is *recognized*/accepted by system *s* that φ is the case"

- Probably we can state that $\overleftarrow{D}_a\, doc-own(a, xpto) \rightarrow [\overrightarrow{D}_a \varphi]\, R_s\, E_{a:owner\text{-}of(xpto)}\, \varphi$

  (where *doc-own(a,xpto)* means "*a* has exhibited the document of ownership of building xpto"), meaning that when *a* shows the certificate of ownership of xpto, then any direct act he will made in that moment will be recognized (or assumed), by the relevant system *s*, as an act playing the role of owner of xpto

- Probably we can even write that $<[a:\top]>\, R_s\, E_{a:owner\text{-}of(xpto)}\, \varphi \rightarrow \overleftarrow{D}_a\, doc-own(a, xpto)$

  sentence that would mean that if it is possible for *a* to make an action that will be recognized as an act in the role of owner of building xpto, for bringing about some state of affairs φ (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 and further work

- We have:

✓ extended the deliberative stit operator to groups of agents;

✓ described the direct and immediate effects of the agent's actions, within the stit framework, both in the achievement sense and in the deliberative sense, without any assumptions on the nature of time and considering the case of a discrete time;

✓ defined a kind of dynamic logic operator that allows us to express what would obtain, if a direct sees to it action is performed, seeing it as a particular case of formulas of the form $[a: \varphi]\psi$, with the following informal meaning: "after *a* choosing φ, ψ is be the case".

- Further research is needed on:

✓ how to implement the *nearly now* requirement;

✓ characterizing the logical properties of the modal operators defined;

✓ integrating these operators with the counts-as, acting in a role and deontic operators, combination that corresponds to the logical setting we think necessary to characterize organizations and other forms of collective acting and interacting, at an adequate abstract level;

✓ illustrating the use of this logical framework within practical and relevant examples.