Collective agency, direct action and dynamic operators

JOSÉ CARMO, *Centre of Exact Sciences and Engineering, University of Madeira, Campus Universitário da Penteada, 9000-390 Funchal, Madeira, Portugal.*
E-mail: jcc@uma.pt

Abstract

We review the stit semantic framework and the main stit operators that have been proposed by Belnap, Perloff and Horty (among others), a theory that has recently attracted the attention of the multiagent community. We discuss the problem of how to model the notion of collective agency, both in the sense of a joint action of a group of agents and as the agency of collective entities, like organisations. We show how we can define the direct and immediate effects of agents' actions, both in the achievement sense and in the deliberative sense, without any assumptions on the nature of time and considering the particular case of a discrete time. Finally, we also show how we can define 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 formulae of the form $[a{:}\ \varphi]\psi$, that have 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").

Keywords: Action, agency, collective agency, direct action, dynamic operators, stit.

1 Introduction

Our work has been to try to characterise those concepts that are essential for the understanding of (organised) collective agency and agent's acting and interacting in general, through the use of modal logic operators.

Particularly important in this context is the study and the characterisation of organisations and other normative multiagent systems, informally seen as entities that are composed by a set of agents and whose behaviour is intended to be governed by norms. At the most basic level, we may see the norms as a way to direct the agency of the different agents, stressing that some acts (in certain circumstances) should be considered as obligatory, permitted or forbidden. Thus, it is only natural that for the characterisation of such systems we try to use and combine deontic operators, like O, P and F (meaning obligation, permission and prohibition) and action operators.

Naturally, the question is which kind of deontic and action logics we should consider. The idea is that we should try to choose the simplest ones that do the job, i.e. that are appropriate to represent and reason about the relevant concepts within the applications we have in mind. In this paper we will concentrate on the action logics, taking into account that we are mainly concerned with a first level of specification of organisations, where we want to characterise and model human and organisations interaction at an abstract level, and where we do not know yet or we do not care about the exact type of actions that can be executed. At that 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

Vol. 18 No. 1,
doi:10.1093/jigpal/jzp072

purposes, what is relevant is that a has brought about some state of affairs φ, and he is responsible for that; or that a is obliged to bring about that φ (being possible that a can do it by different means); etc.[1]

Our approach has been the following: we take advantage of previous contributions of applied modal logic to the representation of social interaction; we confront their expressive power with further concepts relevant for our applications; and, where necessary, we propose additional modalities in order to cope with those concepts.

In their pioneering work, Kanger, Pörn and Lindahl have (precisely) combined deontic and action logics as basic building blocks to describe social interaction and complex normative concepts (see e.g. [27, 28, 30, 32, 33]). Their logics have sufficient expressive power to be able to articulate several distinctions at an appropriate abstract level, mainly by virtue of the modal logic of action they employ. They introduce a relativised modal operator, here generically[2] designated by E_a, where $\mathrm{E}_a\varphi$ means that "agent a brings it about that φ" or "agent a 'sees to it' that φ is the case". Contrary to the dynamic logic[3] operator, that is centred on the specific actions that can be performed, the "sees to it" modal operator relates an agent to the effects of his or her action, omitting details about the specific action that was performed (which seems to be in accordance with the level of abstraction we want for the applications we have in mind).

Central to this "brings it about" concept is the notion of agency and of causation and responsibility, and this approach to the logic of action offers an expressive power rather different from the one provided by others' logics of action, which is particularly useful in some contexts, like those related to the study and representation of social interactions. For instance, using E_a one can express several different positions in which an agent a might be with respect to a certain state of affairs φ, such as $\mathrm{E}_a\varphi$ (did), $\mathrm{E}_a\neg\varphi$ (averted) and $\neg\mathrm{E}_a\varphi\wedge\neg\mathrm{E}_a\neg\varphi$ (remained passive), as well as notions of control of other agents, like $\mathrm{E}_a\mathrm{E}_b\varphi$ (made b do), $\mathrm{E}_a\neg\mathrm{E}_b\varphi$ (made b avoid), etc. Moreover, combining E_a with deontic operators[4] we can then talk about the different normative positions in which one or more agents might be, and use that to express legal concepts and relations like rights, duties, etc., as has been done e.g. in [30].

As a very simple example of some type of analysis that can be made with the previous combination of operators, consider the example, studied in [25], related to the representation of the following hospital regulation: "Patients do not have the right of access to his or her own records". Considering e.g. that $\mathrm{rmr}(x,y)$ denotes "x reads medical record of y", d denotes a doctor and p a patient, then we can express and discriminate different possible alternatives for the formal representation of such regulation: does it mean that $\mathrm{PE}_d\mathrm{rmr}(d,p)\wedge\neg\mathrm{PE}_p\mathrm{rmr}(p,p)$? or does it mean that there is also an obligation of the doctor to see to it that the patient does not have access to his own records, as is expressed

[1] Naturally, for the full specification of an organisation much more must be taken into account.

[2] Other symbols will be used when referring to some specific action operators of this kind.

[3] Dynamic logic was developed within Computer Science, related to the correctness proofs of programs (see [21] for an overview). It uses normal modal operators of the form $[\alpha]$, for α a program, where $[\alpha]\varphi$ means that "(if α is executed) after α, φ is the case". Later it has been generically applied to express the effects of actions, considering other kinds of action terms.

[4] We are here assuming that we can represent personal deontic operators through iterations of impersonal deontic operators and personal action operators (of the previous kind). However, in order to do that we should consider that operator O does not satisfy the RM-rule (if $\vdash\varphi\rightarrow\psi$ then $\vdash \mathrm{O}\varphi\rightarrow\mathrm{O}\psi$), since otherwise we get problems, relative to the transmission of obligations and exclusion of conflicts. (See, for instance, [9, 13] for a discussion on this.)

by $\mathrm{PE}_d \mathrm{rmr}(d,p) \wedge \neg \mathrm{PE}_p \mathrm{rmr}(p,p) \wedge \mathrm{OE}_d \neg \mathrm{E}_p \mathrm{rmr}(p,p)$? 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 formally discriminate different meanings, allowing the user (e.g. the hospital administration) to clarify more precisely what he has in mind with such regulation[5].

The next section of this paper is devoted to the study of this ''sees to it'' operator. We state the logical properties that are usually assigned to this action/agency modal operator and we take a brief overview of the main proposals that have been made for its characterisation, taking particular attention to the ''stit theory'' of Nuel Belnap and Michael Perloff, where an intuitive temporal semantics is provided for such operator. We will describe in detail the stit theory and framework, a theory that has recently attracted the attention of the multiagent community (with proposals of extending the Alternating-time Temporal Logic (ATL) with a *strategic* version of the stit operators, as in [6, 7]), and a theory that we also want to extend in this paper.

Naturally, if we want to describe social interactions we also need to take into account joint actions and collective agency, and section 3 of this paper is devoted to the analysis of such issues.

Two or more agents can jointly act in order to do some tasks. like moving a very heavy table, or to draw up a contract, etc., and the ''sees to it'' / ''brings it about'' action operators can be generalised in order to cover such situations too, as was proposed for instance by Lars Lindahl [30] (although not providing semantics for these operators). 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 $\mathrm{E}_{\{a,b\}}(\mathrm{OE}_a\varphi \wedge \mathrm{OE}_b\psi)$. In section 3 we will discuss such extension of the ''sees to it'' operator, within the stit approach and semantics.

On the other hand, we should distinguish between this kind of *collective agency*, that corresponds to a simultaneous direct action of two or more agents, from the agency of entities that are created by one or (generally) more agents in order to pursue some goals and that are composed or associated with many agents (and can be seen as organised multiagent systems), as it is the case of the organisations.

Such entities (that we may call *collective agents*) also act and interact with the external world (with the other agents), making for instance contracts, and can be subject of obligations and be responsible by their violation / non-fulfilment. However, contrary to what happens with the other agents, a collective agent cannot act, directly. And it is not correct to say that an act of a collective agent corresponds to a joint direct act of all its members. So, how do such collective agents act? What happens is that it acts through the acts of other agents, that act in its name. In general, particular relationships are established between the organisation and other agents, and there are rules that state that when such agents act within the scope of such relationships (within such *role*), such acts count as if it was the proper organisation that has acted.

But an agent can be the holder of different roles within the same organisation, or in different organisations, and can *act by playing different roles*. Moreover, 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 this reason, in [13, 31] we

[5]See [25, 36, 40] for the study of the theory of normative positions, its generation and its usefulness, with a much more detailed analysis of this and other examples.

have extended the previous action 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. It is fundamental to know which acts *count as* acts in a particular role. And, on the other hand, *in reality what we have is agents directly acting*. Thus, the precise question is: in what conditions such direct acts will be recognised as acts in some role by the environment, organisation, society, normative system, etc.?

Therefore, it is important to have the means to discriminate between 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 institutional connection (by the norms applicable e.g. within the relevant organisation). And we may conceive a ''direct'' action operator, for the moment denoted by D, with $D_a\varphi$ meaning that ''agent a has (just) directly brought it about that φ is the case''[6], and continue to use the operator E to express states of affairs that are generically brought about by the agents, in some generic sense.

For instance, in a situation where an agent a steals the canteen of b, when it was in the desert, and in consequence of this, some hours later, b dies of thirst (dehydrated), we may state that $E_a\text{dies}(b)$, but not that $D_a\text{dies}(b)$ (although we can 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 $D_a\text{dies}(b)$. In the applications we have in mind in this paper, we are not interested in specifying those indirect effects that follow from the agents' actions, sometime later, by some causal connections (as in the previous example), but we are particularly interested in being able to specify and distinguish the direct effects of an agent's actions from the institutional, or legal, consequences that may follow from them in some circumstances, by the norms applicable (where, for instance, we want to express that some agent has brought some state of affairs, not because he has directly acted, but because some other agent has directly acted, on his behalf).

The characterisation of such direct action operator, within the stit approach, is the topic of section 4 and it corresponds to one of the items contributing to this paper.

Finally, although the level of abstraction provided by the ''sees to it'' operators seems to be appropriate for the applications we want to model, the fact that these operators are ''static'', providing no resources for (among other things) representing and reasoning about the effects of state change, also limits its applicability (see [37] for a discussion on this). In particular, we would like to be able to state what would obtain if an agent chooses to do some act, allowing us 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.

One of the logic formalisms that was designed to express the effects of actions is dynamic logic, where we can write sentences of the form $[\alpha]\varphi$ (where α is an action term, $[\alpha]$ is a normal modal operator and φ a formula), informally read as follows ''if α is performed, after α, φ is the case''. Thus it is natural that we try to combine these ''sees to it'' action operators

[6] Or that ''φ is the immediate effect of an agent's a direct act'' (a reading in which we will be also particularly interested). Although when we describe these operators informally, the distinction between these two readings is sometimes not clear, when we try to provide a temporal semantics for these operators, as we will do later, the distinction between the reading into the past or related to the future becomes clear.

with the kind of reasoning that is provided by the dynamic logic approach[7], and the last main topic of this paper is how to define an adequate kind of dynamic operators, within the temporal stit semantics. Section 5 is devoted to this.

According to our point of view, the problem of how to adequately combine the static perspective, underlying the ''sees to it'' operators, with a dynamic perspective (that allows us to talk about state changing) has been an open issue for a long time. And we should stress that the proposal we explore here is not the only possibility open for such combination. For instance, in a very recent (and very interesting) paper [38], a completely different approach has been proposed by Marek Sergot, based on extensions of labelled transition systems. Although a detailed comparison between the two approaches is outside the scope of the present paper, in the last section (before the concluding remarks) we will make a brief overview of Sergot's proposal and a first comparison with our approach here.

We will use: a, b, ... for (names of) agents; φ, ψ, ... for formulae; p, q, ... for atomic sentences; and G, G′, ... for groups of agents. For the propositional (Boolean) connectives we will use: $\neg$, $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$; $\top$ will denote a tautology and $\perp$ a contradiction. The following precedence rules will be assumed: 1st) unary operators; 2nd) $\wedge$; 3rd) $\vee$; 4th) the other binary operators.

2 *Sees to it* logics

In this section we make a brief overview of the ''sees to it'' logics, with an emphasis on the stit theory.

2.1 The standard logical properties assigned to the ''sees to it'' operator

Although the formal properties assigned to the action operator E_a may vary among the different authors, with the main exception of Brian Chellas [14, 16], that proposes for E_a a normal modal logic of type KT (using Chellas's classification [15]), E_a is usually considered a non-normal modality satisfying the rule (schema)

$$\text{RE:}\quad \text{if } \vdash \varphi \leftrightarrow \psi \text{ then } \vdash E_a\varphi \leftrightarrow E_a\psi \quad \text{(replacement of logical equivalents rule)}$$

and including the schemas

$$\begin{array}{ll} \text{T:} & E_a\varphi \rightarrow \varphi \\ \text{C:} & (E_a\varphi \wedge E_a\psi) \rightarrow E_a(\varphi \wedge \psi) \\ \text{No:} & \neg E_a \top \end{array}$$

(That is, it is usually considered that the set of theorems, or of valid formulae, of a ''sees to it'' modal logic includes all instances of T, C and No and is closed under RE).

The T-schema captures the intuition that if agent a brings it about that φ, then φ is indeed the case; that is, E_a is a ''success'' operator. (Another way of arguing for the necessity of schema T is to say, following Chellas [14], that ''a can be held responsible for its being the

[7]Note that we are not saying that we are going to use dynamic logic or defending its use for multi-agent systems specification (for a brief discussion on the benefits of using stit formalisms, instead of dynamic logic, as a basis for logics for multi-agent systems see e.g. [8]). What we say is that we intend to incorporate within the stit formalisms the kind of hypothetical reasoning that is provided by the dynamic logic approach.

case that φ, only if it is the case that φ''.) Schema No 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 truth of $E_a\varphi$ must imply that the action of a was necessary to get φ; no agent can meaningfully bring about what is logically true, or more generally, what was unavoidable.

Note that we do not want the logic of the operator E_a to be closed under logical consequence, i.e. E_a should *not* satisfy the rule (schema)

$$\text{RM:} \quad \text{if } \vdash \varphi \rightarrow \psi \text{ then } \vdash E_a\varphi \rightarrow E_a\psi$$

(otherwise, since we have imposed the schema No, we would have that $\vdash \neg E_a\varphi$, for all a and φ).

With respect to the semantic characterisation of this operator, we find different proposals[8], running from a combination of two accessibility relations, as Kanger and Pörn do, to the use of variants of the minimal models (popularized by [15]), as in [17, 18, 34]. For instance, and just as an illustration:

- Pörn uses, in his models M, two disjoint accessibility relations (associated to each agent a), $R1_a$ and $R2_a$, the first reflexive and the second serial and irreflexive, and defines

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

  where, informally, $R1_a$ relates a world w with the worlds v where the agent a 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;
- And in [34] the models just include, for each agent a, a function f_a: $\wp(W) \rightarrow \wp(W)$, 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), and defines

$$\mathsf{M} \models_w E_a\varphi \text{ iff } w \in f_a(||\varphi||) \quad (\text{where } ||\varphi|| = \{w\colon \mathsf{M} \models_w \varphi\})$$

  (The minimal models considered in [17, 18] are more complex.)

Besides the previous semantic approaches, we have also approaches based on temporal frameworks, that we will present next, where we find an attempt to give a (less abstract and) more informative and intuitive semantics to these action operators. Particularly important in this area is the ''stit theory'', initiated by Belnap and Perloff [1, 2].

2.2 The achievement stit operator (astit)

The stit's models use histories, moments and agents' choices. The models are based on tree-like structures; at each moment the past is linear and there exist open future branches. We may see a history as a complete branch in the tree and define truth relative to a pair formed by a history and a moment (belonging to such history). The central element in their analysis is the notion of agent's choice: all histories passing through a same moment are partitioned in a set of equivalent classes, corresponding to the choices (or actions) that are open to the agent at that moment.

[8]For overviews of the main approaches see e.g. [17, 18, 34], and particularly [22], where Kanger, Pörn, von Wright, Belnap and many others' approaches on action and agency are analysed.

Within this framework various *stit* ("sees to it") operators have been proposed. In what follows we will present the two main stit operators: the first *stit* operator, now called the *achievement stit* (*astit*, for short) in order to distinguish it from the other stit operator, called *deliberative*[9] *stit* (*dstit*, for short). We will start by the former.

The main idea underlying this concept is the following: an agent a sees to it that φ, which is expressed through a sentence of the form [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). Formally, we may describe the stit semantics as follows.

A *tree-like frame* is a pair (Mom, $\leq$) where Mom is a nonempty set, whose elements $m, m', \ldots$ are called *moments*, and $\leq$ is a partial ordering on Mom representing the temporal earlier/later relation among moments (m is earlier than m', $m<m'$, if $m \leq m'$ and $m \neq m'$). The ordering $\leq$ is subject to the following condition[10]:

- *No downward branching* (the past is linear / future branching only): $\forall_m \forall_{m'} \forall_{m''} (m' \leq m \wedge m'' \leq m \rightarrow m' \leq m'' \vee m'' \leq m')$

A *history* in a tree-like frame (Mom, $\leq$) is a branch of the "tree", i.e., a maximal $\leq$-chain h. A history *passing through* m is a history to which m belongs. H denotes the set of histories and H_m the set of histories passing through m (i.e. $\mathrm{H}_m = \{h(\in \mathrm{H}): m \in h\}$). We say that two histories, h and h', are *undivided* at a moment m, if it exists a moment $m' \in h \cap h'$ such that $m<m'$; and that h and h' *split* at m, if h and h' pass through m and h and h' are not undivided at m.

In order to define the achievement stit operator, it is also considered in the stit frames (see e.g. [2], or [4, pp 35]) that the tree can be partitioned horizontally into *instants* (of time), assuming that all branches of the considered tree have a unique temporal order[11], being possible to define an isomorphism[12] between any two histories (and the moments belonging to the same instant are said *co-instantial*). In order to simplify the presentation[13], given any

[9]The term 'deliberative' does not mean that the indicated outcome was intended. No intentions are presupposed in the stit operators, as it is clear from this passage from [4, pp 33] (see also [24, pp 12]) "Now what is it for one of these agents to act, or choose, in this way? We idealize by ignoring any intentional components involved in the concept of action, by ignoring vagueness and probability, and also by treating actions as instantaneous." The term deliberative has to do with the fact that "it is usually natural to take the complement of a deliberative stit as future tensed" (see [4, pp 37], or e.g. [24, pp 16]).

[10]In some works, like [44], the following *historical connection* constraint is also considered: $\forall_m \forall_{m'} \exists_{m''} (m'' \leq m \wedge m'' \leq m')$. Although this condition has no impact on the valid stit formulae, if we have both *no downward branching* and *historical connection*, then a tree-like frame really looks like a single "tree".

[11]See [42] for a discussion on this assumption.

[12]Naturally, if a moment m belongs to two histories, we assume that such isomorphism applies m into itself.

[13]We note that it is possible to redefine these tree-like frames, as Chellas has done in [16], as structures (T, $\leq$, S, H), where (T, $\leq$) represents the *time* structure, and can be any linear ordered set (so that we can speak unambiguously of one time, or instant of time, being earlier than or later than another), S is a nonempty set of elements that represent the *possible moments* or *states of affairs* (elements that for any relevant purpose can be let unspecified), and H is a nonempty set of *histories*, with a history a function from T to S (where $h(t)$ represents the moment in history h at instant t), satisfying the conditions: i) (future branching only): $\forall_t \forall_h \forall_{h'} (h(t) = h'(t) \rightarrow \forall_{t' \leq t}\, h(t') = h'(t')$ and ii) (moments are all distinct): $\forall_h \forall_{h'} \forall_t \forall_{t'} (h(t) = h'(t') \rightarrow t = t')$. Condition ii) is not considered in [16], but it makes these frames exactly as the tree-like frames above, where a same moment cannot appear twice in a history, neither can it appear in two different histories, unless it corresponds to the same instant of time, having then, by i), the same past in both histories. (If we want to consider the *historical connection* constraint, then we should also impose that $\forall_h \forall_{h'} \exists_t h(t) = h'(t)$.) The set Mom of the *moments* in the tree may then be defined as Mom $= \cup_{h \in H} h(\mathrm{T})$, and we may still write $m \in h$ meaning that $\exists_t (h(t) = m)$. In fact, for a moment m in the tree, there exists one and only one time instant t such that $h(t) = m$, and if we denote such time instant by i_m,

moment m and any history h, we write $m@h$ to denote the unique moment belonging to h that is *co-instantial* with the moment m.

A *stit frame* is a structure $\mathsf{F} = (\text{Mom}, \leq, \text{H}, \text{Ag}, \text{C})$ obtained by enriching a tree-like frame (Mom,$\leq$) with two components: a nonempty set – the set Ag of all *agents* (names), and a choice function C:MomxAg$\rightarrow \wp(\wp(\text{H}))$, which is a mapping that applies each moment m in the tree and each agent a into a partition of the set H_m.

The elements of $C(m,a)$ correspond to the *actions*, or *choices*, available to a at m. We use $C_{m,a}(h)$ (only defined when h passes through m) to denote the particular action or choice from $C(m,a)$ that contains history h (the one that was selected by a, at the moment m, if such history is followed); and write $C_{m,a}(h,h')$ to denote that h and h' pass through m and belong to the same choice set at m (thus, $h' \in C_{m,a}(h)$). When $\#C(m,a)=1$ (i.e. $C(m,a)=\{\text{H}_m\}$), we say that a has a *vacuous choice* at moment m.

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

- *No choice between undivided histories*: "(for any moment m, any agent a, and any two histories h and h'), if h and h' are undivided at m, then $C_{m,a}(h,h')$";
- *Something happens*: "for every way of selecting one possible choice for each agent from his or her set of choices, the intersection of all possible choices selected must contain at least one history", i.e., more precisely, "for any moment m and for any map s_m:Ag$\rightarrow \wp(\text{H})$, such that $s_m(a) \in C(m,a)$, we have that $(\bigcap_{a\in Ag} s_m(a)) \neq \varnothing$".

According to the *something happens* constraint, each agent can choose an alternative in her/his choice independently of the alternatives chosen by all the other agents (at the same moment). In particular, at a moment m, no agent can prevent another agent from choosing any of her/his alternatives (at m). We will return to this condition later.

Within this temporal framework, the truth value of a sentence is evaluated with respect to history-moment pairs (h,m), with h a history passing through m. We write $\mathsf{M} \models_{h,m} \varphi$, read "formula φ is *true* (in model M) *at* (h,m)", meaning that φ is true at the moment m, assuming that history h is followed (or according to history h). A formula φ is *true in* a model M, written $\mathsf{M} \models \varphi$, if it is true in M at all pairs (h,m) (such that $h \in \text{H}_m$) and φ is *valid*, written $\models \varphi$, if it is true in all models. As usual, we use $\not\models$ to deny $\models$.

The models take the form $\mathsf{M}=(\mathsf{F}, v)$, where[14] the *valuation* v applies each atomic sentence p into a set of pairs (h,m), with $h \in \text{H}_m$ (the pairs at which, intuitively, p is thought of as true). Although Belnap does not impose this condition[15], we will assume that valuations satisfy the following condition (where we use the terminology of [16])

- *Extensionality condition for atomic sentences*:
  $\forall_h \forall_{h'} \forall_m \ (m \in h \cap h' \rightarrow ((h,m) \in v(p) \text{ iff } (h',m) \in v(p)))$

since we think that the truth values of the atomic sentences should only depend on the moment where they are evaluated, and not on its past, or possible future.

then the abbreviation above $m@h'$ simply corresponds to the moment $h'(i_m)$ (in another way, if $m = h(t)$, then $m@h' = h'(t)$). Although, in our opinion, the semantics of the *astit* operator (and of the Chellas's action operator, to be introduced later) can be presented in a more elegant and simpler way within this alternative framework, we will keep the standard stit semantics, since for the other stit operators we do not need to assume that all branches of the tree have a unique temporal order.

[14]We follow the practice of using the same symbols for agents and agent symbols.

[15]But he also does not refuse it (see [4, pp 31, footnote 4]).

Within this rich temporal framework, we evaluate the truth-value of the atomic sentences and of the usual truth-functional connectives, as expected, without changing the index (h, m) of the evaluation (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.), and several kinds of modal operators can be defined, changing the moment index m through the history h, changing the history index h and keeping fixed the moment m, or changing both indices.

For instance, the usual linear temporal operators can be defined changing the moment index m through the history h, that is kept fixed. As an example, the *until* operator[16] U^{ee} is defined as follows

$$\mathsf{M} \models_{h,m} \varphi U^{ee} \psi \text{ iff there exists a moment } m^* \in h \text{ such that } m^* > m, \mathsf{M} \models_{h,m^*} \psi \text{ and, for any } m < m' < m^*, \mathsf{M} \models_{h,m'} \varphi$$

Note that with this operator we can get the other future linear operators. For instance, the operator W, that we will consider later, proposed by Mark Brown (in [10]) to capture the notion of *for a while in the immediate future*, can be defined as follows $W\varphi =_{\text{def}} \varphi \mathrm{U}^{ee} \varphi$. And, if we assume time to be discrete, isomorphic to the set of integers or to the set of natural numbers, then we can define the next time operator X by[17] $X\varphi =_{\text{def}} \bot \mathrm{U}^{ee} \varphi$. (Observe that, within such time, $W\varphi$ and $X\varphi$ are equivalent.) Naturally, parallel linear operators can be defined respecting the past, like the since operator:

$$\mathsf{M} \models_{h,m} \varphi\ \mathrm{S}^{ee} \psi \text{ iff there exists a moment } m^* \in h \text{ such that } m^* < m, \mathsf{M} \models_{h,m^*} \psi \text{ and, for any } m^* < m' < m,\ \mathsf{M} \models_{h,m'} \varphi$$

and, within discrete time, the definition $Y\varphi =_{\text{def}} \bot \mathrm{S}^{ee} \varphi$ gives now the previous time operator [18] Y.

On the other hand, we can define necessity operators changing the history index h and keeping fixed the moment[19] m, like in the definition of the following *historical necessity* operator (“settled true at m”)

$$\mathsf{M} \models_{h,m} \Box \varphi \text{ iff } \mathsf{M} \models_{h',m} \varphi \text{ for every history } h' \text{ passing through } m \text{ (i.e. } h' \in H_m)$$

Finally, the truth condition for the more complex *astit* action operator, is as follows[20]:

$\mathsf{M} \models_{h,m} [a \textit{ astit: } \varphi]$ 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 $h' \in \mathrm{H}_{m^*}$

[16]Following [6], we use the superscript ‘ee’ to denote that this is the version of ‘the until operator’ where φ is not required to hold for the current moment, nor for the moment where ψ holds (*both* moments are *excluded*).

[17]Using $m+1@h$ to refer to the moment after m, in history h, we can directly define the semantics of X as follows: $\mathsf{M} \models_{h,m} X\varphi$ iff $\mathsf{M} \models_{h,\ m+1@h} \varphi$.

[18]Using $m-1$ to refer to the moment before m, we can directly define the semantics of Y as follows: $\mathsf{M} \models_{h,m} Y\varphi$ iff it exists $m-1$ and $\mathsf{M} \models_{h,m-1} \varphi$, within natural time, and $\mathsf{M} \models_{h,m} Y\varphi$ iff $\mathsf{M} \models_{h,m-1} \varphi$, within integer time.

[19]We can also define necessity operators where both indices change, as in the following definition of an *always* operator: $\mathsf{M} \models_{h,m} \boxed{\Box} \varphi$ iff $\mathsf{M} \models_{h',m'} \varphi$ for every m' and every $h' \in H_{m'}$.

[20]Note that this definition does not cover all cases of “seeing to it” (“bringing it about”). Namely, it does not cover the cases where an agent does an action that although not guaranteeing a possible outcome, may lead to that outcome, which would not obtain if the agent had not acted as he did. Consider e.g. that an agent a dies by executing the action of playing Russian roulette. In such case the agent would be certainly responsible for his own death, although the choice made by a of playing Russian roulette does not guarantee that a will die. We may say that the *astit* operator expresses a kind of “strong” seeing to it concept.

The choice points corresponding to the pairs (h, m^*) where both the previous positive condition (1) and the negative condition (2) are satisfied, are referred by Belnap as ‘‘witnesses’’ (for the truth of $\mathrm{M} \models_{h,m} [a \textit{ astit: } \varphi]$). It can be shown that such witnesses are unique (see e.g. [16]).

The set of valid formulae that we get for the *astit* operator, according to this semantic definition, corresponds to a non-normal modal logic, obeying the RE-rule and including the schemas T, C and No, plus (among others[21]) the schema

4: $[a \textit{ astit: } \varphi] \rightarrow [a \textit{ astit: } [a \textit{ astit: } \varphi]]$

On the other hand, with respect to iterations of these action operators, related to different agents, we have (it is valid) the following schema (if $a = b$, Q follows from T)

Q: $[a \textit{ astit: } [b \textit{ astit: } \varphi]] \rightarrow [a \textit{ astit: } \varphi]$

that expresses a kind of principle of responsibility: if agent a ‘‘sees to it’’ that agent b ‘‘sees to it’’ that something is the case, then *ipso facto* a ‘‘sees to it’’ (too)[22].

As [16] refers, although the schema Q is valid even if we do not assume the *something happens* condition, if we assume such a condition, then the validity of Q becomes trivial, since the following schema then also becomes valid

NoControl: $\neg[a \textit{ astit}: [b \textit{ astit}: \varphi]]$, whenever the agents a and b are different

If we want to model humans’ interaction, surely there exist cases where an agent is (also) responsible for some outcome, because he or she has forced another agent to directly produce such an outcome. The *something happens* condition does not allow us to express such situations. For this reason, Chellas (in [16]) states that ‘‘the correctness of the *something happens* condition must be doubted’’.

Within applications where we are not interested in expressing the control that an agent can exercise on other agents, and where we are only interested in expressing the effects of the agents’ direct actions and its consequences (possibly on other agents), we may assume the *something happens* condition. Otherwise, we should abandon it, which is compatible with keeping the rest of the stit semantics. Note that if we do not impose the *something happens* condition, when a history h is followed at a moment m, we still see that all agents have made some choice (forced or not), since h always belongs to some set in $C(m, a)$, for every agent a.

2.3 The Chellas ‘‘sees to it’’ operator

The Chellas ‘‘sees to it’’ operator (denoted in [16] by) Δ_a, for a an agent, tries to capture a conception of agency as dependent on the past, where the states of affairs under an agent’s control are limited to those that are possible outcomes of the way the world has been up to (but not including) a moment of agency.

This operator can be defined through a temporal framework similar to the previous one, replacing (in the stit frames) the choice function C by a family – $(\mathrm{R}_{m,a} : m \in \mathrm{Mom}$ and $a \in \mathrm{Ag})$

[21] Namely, we omit here any discussion about the relationships between *refraining from refraining from seeing to*, and *seeing to* (doing).

[22] Chellas says that Q encapsulates the legal maxim *Qui facit per alium facit per se* (acknowledging Noyes Leech for this remark).

- where each $R_{m,a}$ is intended to describe (to relate) the histories that are *instigative alternatives*, i.e. "under the control of – or responsive to the actions of" a at m. More precisely, each $R_{m,a}$ is a binary relation on the set (that we denote by) $H_{<m} = \{h: \{m^*: m^* < m\} \subseteq h\}$, i.e. the set of histories that contain all the past of the moment m. (We will write $R_{m,a}(h, h')$ for $(h, h') \in R_{m,a}$). It is only imposed that each relation $R_{m,a}$ is reflexive[23].

The truth condition for the operator Δ_a is defined as follows[24]:

$$\mathsf{M} \models_{h,m} \Delta_a \varphi \text{ iff } \mathsf{M} \models_{h', m@h'} \varphi \text{ for every history } h' \text{ such that } R_{m,a}(h, h')$$

We get for each operator Δ_a a normal modal logic satisfying (among others) the T-schema $(\Delta_a \varphi \to \varphi)$, and iterations of these operators satisfy (not trivially[25]) the schema Q $(\Delta_a \Delta_b \varphi \to \Delta_a \varphi)$. Although the schema 4 $(\Delta_a \varphi \to \Delta_a \Delta_a \varphi)$ is not valid, (as Chellas observes) we get it if we also impose that the relations $R_{m,a}$ are transitive.

2.4 The deliberative stit operator (dstit) and the Chellas stit operator (cstit)

In 1995, Horty and Belnap [23] proposed a new stit operator, which they called the *deliberative* stit (*dstit* for short), defined as follows (where M is a stit model, but where it is not necessary that all branches of the tree have a unique temporal order):

$\mathsf{M} \models_{h,m} [a \textit{ dstit: } \varphi]$ iff
(1) $\mathsf{M} \models_{h',m} \varphi$ for every history h' such that $C_{m,a}(h, h')$ (i.e. $h' \in C_{m,a}(h)$), and
(2) $\mathsf{M} \not\models_{h',m} \varphi$ for some history h' that pass through m (i.e., such that $h' \in H_m$)

The deliberative stit sentences concern the agent's present choices / actions, whereas the achievement stit sentences refer to the agent's past choices.

Like the *astit* operator, the *dstit* operator, so defined, is also a non-normal modal operator, obeying the RE-rule and including (among others) the following (valid schemas): T, C, No and 4. Also, like the *astit* operator, the *dstit* satisfies the NoControl schema, because of the *something happens* condition, and without such condition, the Q-schema would still be valid.

If we consider the extensionality condition for the atomic sentences (as we did), then, if φ is any formula without modal operators (i.e. composed only of propositional symbols and truth-functional connectives), or including only modal operators that satisfy the extensionality condition, as past operators or $\Box$. then $\models \neg[a \textit{ dstit}: \varphi]$ (since, in such case, $\models \varphi \to \Box\varphi$). We can say that an agent can deliberatively sees to it only sentences about the future.

On the other hand, we can describe $[a \textit{ dstit: } \varphi]$ using the necessity operator $\Box$ and an action operator that expresses only the positive condition (1) above. In fact, defining

$$\mathsf{M} \models_{h,m} [\ a \textit{ cstit}: \varphi] \text{ iff } \mathsf{M} \models_{h',m} \varphi \text{ for every history } h' \in C_{m,a}(h)$$

[23]Intuitively, we think we should impose that if h and h' belong to $H_{<m}$ and both pass through m, or both pass through a same moment co-instantial with m, then $R_{m,a}(h, h')$ (which implies the reflexivity of $R_{m,a}$).

[24]Informally, we may see Pörn's relation $R1_a$ having a mean analogous to Chellas's $R_{m,a}$, in the sense that, if we identify an abstract world w with a moment m plus their past (or, in another way, with any pair (h, m), with h passing through m), then we may consider that $(h, m) = w\ R1_a v = (h', m')$ iff $h\ R_{m,a} h'$ and $m' = m@h'$. In this sense, Chellas's definition is like Pörn's definition, but omitting the negative condition.

[25]Note that Chellas does not consider any constraint equivalent to the *something happens* condition.

we have that (where $\diamond$ denotes the dual operator of $\Box$)

$$\models [a\ dstit\colon\ \varphi] \leftrightarrow [a\ cstit\colon\ \varphi] \wedge \diamond\neg\varphi$$

In [23], the *cstit* operator is described as the "Chellas stit" because it is said that its definition resembles the Chellas sees to it operator. However, as it is referred in [22], there is a significant difference between them, since in the *cstit*'s definition the alternative histories h' under analysis must pass through the moment m, and so must have the same present and past that h has at m, whereas in Chellas the alternative histories h' under consideration must only have the same past that h has at m.

As it is referred in [6], one of the virtues of these latter stit operators is that "unlike most (if not all) other logical formalisms, they can express that a choice or action is actually performed / taken / executed by an agent".

The distinction between the stit sentences concerning the agent's past choices (actions), and the stit sentences concerning the agent's present choices is important in many respects. In particular, it may help in clarifying the meaning that we want to give to iterations of deontic operators and "sees to it" action operators. For instance, various authors (including ourselves), when considering a semantic abstracting from temporal issues, usually informally read $\mathrm{E}_a\varphi$ in the past sense (referring to the agent's past choices/actions, like in the achievement stit), but when writing $\mathrm{OE}_a\varphi$ are concerned with the agent's present choices/actions, intuitively meaning that the agent a is under an obligation to "do" φ[26], and not that it is obligatory that a have (just) brought it about that φ. These two different types of obligations become clearly expressed by the iterations[27] $\mathrm{O}[a\ dstit\colon\ \varphi]$ and $\mathrm{O}[a\ astit\colon\ \varphi]$.

2.5 Brief comparison between the stit logic and some multiagent logics

As we indicated in the introduction, the "stit theory" has recently attracted the attention of the multi-agent community, and there already exist some works comparing the stit logic of agency with ATL, like [6, 7, 43], as well as works presenting a translation of Coalition Logic[28] to Chellas' stit logic, like [5]. Thus, for the sake of brevity, we just refer the reader to them, and omit here any comparison between these logics. On the other hand, we delay to section 6, a comparison with the framework (very recently) proposed by Marek Sergot [38, 39], for describing and analysing norm-governed multiagent systems.

Other multiagent frameworks that have also some similarities with stit theory, are those that have been proposed in some of Munindar P. Singh's papers, namely [41].

The model proposed in [41] is based on a set of moments with a strict partial order which denote temporal precedence. Time may be discrete, or not. Although the past may be taken as linear[29], time may branch into the future in order to model agents' choices. Each basic action has exactly one agent performing it, and multiple agents can act simultaneously. Each moment is associated with a possible state of the world, which is identified with the

[26] In which case it is not correct to say that there is a violation when $OE_a\varphi \wedge \neg E_a\varphi$ is the case.

[27] And a violation of any of these two obligations can be correctly expressed as a conjunction of the obligation and the negation of the respective action formula (obliged). Naturally, in reality, in the current moment (without fixing the history that it will be followed), we can only recognize a violation of the latter kind of obligations ($\mathrm{O}[a\ astit\colon\ \varphi]$).

[28] For a comparison between ATL and coalition logic (and other multiagent logics) see, for instance, [19].

[29] For reasons that we will not discuss here, in [41] it is not imposed that the past is linear, although we might assume that.

atomic conditions or propositions that hold at that moment. Each agent influences the future by acting. A *scenario* (S_m) at a moment m is like a history h passing through m (but not considering the past of that moment, i.e. h-$\{m' : m' < m\}$), and different scenarios at m correspond to different combinations of actions by the various agents.

The main difference with respect to stit theory concerns the treatment of actions. In stit, actions are treated as instantaneous and no symbols are introduced in the formal language to denote actions; and in [41] (even basic) actions take time and can be of arbitrary duration, and we can refer to them in the language[30], being the action terms built from the basic action symbols as in the regular programs. In the semantics, to each action α and each agent a is associated a set of periods of the form $[S;m, m']$ (with $m \leq m'$) meaning that a is performing α from moment m to moment m' in the scenario S.

The qualitative temporal language proposed, *TAB*, is based on[31] CTL* (including the future linear until operator U and the quantification over scenarios, that correspond to our necessity operator $\Box$), extended with regular programs, quantification over basic actions, and with three different kinds of dynamic logic operators: $a[\alpha]\varphi$, that holds on a scenario S and a moment m on it, iff, if a is performing α at m, then φ holds at some moment after m (in that scenario) while the given instance of α is being performed; $a < \alpha > \varphi$, that is not a formal dual of $a[\alpha]\varphi$, that is true at S,m, iff a is performing α at m and φ holds at some moment after m (in that scenario) while the given instance of α is being performed; and a stronger $a| < \alpha > |\varphi$, that is true at S,m, if a is performing α at m and φ holds in some initial subperiod of the remaining period over which α is performed.

Although the treatment of actions is different (which has also to do with the applications they have in mind), the temporal framework underlying the two approaches have clear similarities. We also note that later on we are going to extend stit with a kind of dynamic operators. and by different but comparable reasons, we will also need to consider two dynamic operators that are not formal duals.

3 Collective agency

3.1 Joint action

Two or more agents can jointly act in order to do some task, and the "sees to it" action operators can (and should) be generalized in order to cover also such situations.

Suppose that G denotes a group (a nonempty finite set) of agents. Our informal idea regarding this "*collective agency / 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. It is this kind of joint action that we want to capture here.

[30]For an attempt to introduce action names and to incorporate actions whose completion requires an interval of time within stit framework, see e.g. [11].

[31]The distinction in CTL* (that we think it is unnecessary) between moment/state formulae and scenario/path formulae is kept in [41], being the truth value of the former analyzed with respect to a moment and the latter with respect to a moment and a scenario containing that moment. It is also considered a "reality" operator R, corresponding the truth evaluation of Rφ, at a moment m, to the evaluation of the truth of φ at m according to a scenario that is assigned to m in the model, and that is interpreted as being the real scenario at m.

Both the achievement and the deliberative stit operators can be extended in order to cover also this kind of collective agency. We will extend below the simpler deliberative stit operator, but we think that the same ideas also apply to the achievement stit operator.

We think that such extension should satisfy the following two general requirements (where we use *stit* to refer to both the achievement and the deliberative stit operators)[32]:

- *Single group*: if G={a}, then the schema [G *stit:* φ] ↔ [*a stit:* φ] should be valid
- *No group monotonicity*: [G *stit:* φ] → [G′ *stit:* φ], with G⊆G′, should *not* be a valid schema

The *single group* condition is an obvious requirement. If we assume *group monotonicity* then, whenever a group G sees to it that φ, the set of all the agents also sees to it that φ, and this is against the idea of joint action that we are trying to capture here[33]: the agents not in G have nothing to do with the "doing of φ", and cannot also be made responsible for φ being the case.

A first possible obvious extension of the *dstit* operator, is (where M is a stit model):

$\mathsf{M}\models_{h,m}$[G *dstit:* φ] iff
(1) $\mathsf{M}\models_{h',m}\varphi$ for every h' such that $C_{m,a}(h,h')$ for every a in G (i.e. $h'\in\bigcap_{a\in G}C_{m,a}(h)$), and
(2) $\mathsf{M}\not\models_{h',m}\varphi$ for some history h' that pass through m

that corresponds to the *plain join stit* discussed in [4][34].

However, the previous proposal does not serve our purpose, since it validates *group monotonicity*. The problem is on the negative (or counterfactual) condition (2), that it is not strong enough to guarantee that the action of *each* of the agents in G was necessary to obtain φ. In order to guarantee this, we propose to define[35]:

$\mathsf{M}\models_{h,m}$[G *dstit:* φ] iff
(1) $\mathsf{M}\models_{h',m}\varphi$ for every history h' such that $h'\in\bigcap_{a\in G}C_{m,a}(h)$, and

[32]What follows does not apply to the *cstit* operator. As a matter of fact, the natural way to extend *cstit* to groups of agents is to define (as in [24, pp 32]) $\mathsf{M}\models_{h,m}$[G *cstit:* φ] iff $\mathsf{M}\models_{h',m}\varphi$ for every $h'\in\bigcap_{a\in G}C_{m,a}(h)$, and with this definition we satisfy *group monotonicity* requirement. But the *cstit* operator (although very useful) does not express the concept of agency/responsibility we are looking for, since it does not imposes a negative condition. (Nevertheless, we note that, in part for its simplicity, *cstit* is used by various authors in order to express individual or group agency; e.g. the definition of "Group G 'sees to it' that φ" in [29], although defined within a standard possible worlds approach, instead of within stit branching-time models, is of this *cstit* type.)

[33]This distinguishes clearly this approach from others' approaches that have also extended the stit operators with groups of agents, namely those that intend to define a *strategic* version of stit, like [6], that considers sentences of the form [G]φ, read "agents in G strategically see to it that φ", and that satisfy ⊨[G]φ→[G′]φ, for G⊆G′. This comment is not intended as a criticism to those approaches. What happens is that the concept of group agency they try to capture is different from ours. If a group G has a strategy that can ensure the truth of φ, then any bigger group also has a strategy that can ensure the truth of φ. But, from the fact that a group G is responsible for φ being the case, we should not conclude that any bigger group is also responsible for that.

[34]The problem of joint agency, within stit, is analysed in [4, chapter 10] (based on Belnap and Perloff's paper of 1993 [3]). Although the analysis there is directed for the achievement stit, it is also considered that the ideas work equally for the deliberative stit (see [4, pp 271]). Two main proposals are discussed there. The above *plain join stit*, denoted by [G *stit*: φ], and the so called *strict joint stit*, that is defined as follows: [G *sstit*: φ] ↔ [G *stit*: φ] $\wedge\forall_{G'}(\varnothing\neq G'\subset G\rightarrow\neg$[G′ *stit*: φ]).

[35]If G={a}, then we get $\bigcap_{b\in G-\{a\}}C_{m,b}(h)=\bigcap_{b\in\varnothing}C_{m,b}(h)$, and we assume that $\bigcap_{b\in\varnothing}C_{m,b}(h)$= H. The reference in (2) to $h'\in\mathrm{H}_m$ it is only necessary when G is singular (and so G−{a}= ∅). When this is not the case, $h'\in\mathrm{H}_m$ follows from $h'\in\bigcap_{b\in G-\{a\}}C_{m,b}(h)$.

(2) for every a in G, there exists some history h' such that h' pass through m (i.e. $h' \in H_m$) and $h' \in \cap_{b \in G\text{-}\{a\}} C_{m,b}(h)$ and $\mathsf{M} \not\models_{h',m} \varphi$

proposal that can be seen to correspond to the *strict join stit* discussed in [4] (see appendix).

By (1), the history h' provided by (2), for each a in G, must be such that $h' \notin C_{m,a}(h)$. Thus, according to (2), 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).

It is easy to see that, according to the proposed definition, the *single group* condition is satisfied (i.e. $\models [\{a\}$ *dstit:* $\varphi] \leftrightarrow [a$ *dstit:* $\varphi]$), and regarding *no group monotonicity*, we have the stronger result:

- *Group anti-monotonicity*: $\models$[G *dstit*: φ] $\rightarrow \neg$[G′ *dstit*: φ], whenever G ⊂ G′
  (Proof: Suppose, by absurd, that (i) $\mathsf{M} \models_{h,m}$[G *dstit:* φ], (ii) $\mathsf{M} \models_{h,m}$[G′ *dstit:* φ] and (iii) G⊂G′. Let $a \in$G′-G. By the negative condition of (ii), exists $h' \in \cap_{b \in G' - \{a\}} C_{m,b}(h)$ such that $\mathsf{M} \not\models_{h',m} \varphi$. But then $h' \in \cap_{b \in G} C_{m,b}(h)$, and, by the positive condition of (i), $\mathsf{M} \models_{h',m} \varphi$, getting a contradiction.)

Note that this result does not mean that we cannot have a situation where different groups of agents may ''see to it'' the same outcome φ, but in such case neither of the groups is (strictly) contained in the other.

Before proceeding, we should note also that even if we do not assume the *something happens* condition, when we assert [G *dstit:* φ] (according to the semantic definition above) we are considering that the action of each member of G, that jointly lead to the outcome φ, was not forced by the actions of the other members of G (since each agent in G could act differently, while all the other members acted as they did)[36]. Thus, [G *dstit:* φ] means that the agents in G are jointly cooperating to bring about φ (and such cooperation was not forced, although possibly also not intended, occurring by chance).

Each [G *dstit:* _] modal operator is non-normal, obeying the RE-rule and including (among others) the (valid) schemas: T, C, No, 4 and (assuming the *something happens* condition) NoControl[37].

Just as an illustration, let us see how this definition of collective action behaves in two simple concrete examples referred at [30, pp 222]. Case i): suppose four grams of poison is the minimum quantity sufficient to kill a person c; a and b each give c two grams of poison at the same time; and φ is the sentence ''c dies''. In this case, both [30] and this approach would represent this as a case where the two agents a and b jointly have brought about φ, without being the case that any of them has brought it about by himself. Case ii): suppose now that a and b each simultaneously give four grams of poison to c. Then, according to [30], both a, b and *{a,b}* have brought it about φ, whereas, according to our approach here, both a and b have brought it about φ, but *{a,b}* didn't (since the join of the actions of the two agents was not necessary to produce the result).

Clearly, there exist other types of real or ''academic'' cases that our approach does not represent so well. The notion of collective agency is very complex and there are a multitude

[36] And this is reflected in the fact that we can prove the validity of ¬[G *dstit*: [G′ *dstit*: φ]], when one of the sets G or G′ is strictly contained in the other, without using *something happens*. (See appendix.)

[37] The only schemas whose validity is not obvious, are 4 and NoControl. Their proofs (that extend the proofs in [16] for similar results for single agents and the *astit* operator) can be seen at the appendix.

of different situations where we could say that a group of agents is responsible for something being the case. For some cases, our notion of joint action may be considered too strong, and it is possible to conceive definitions that satisfy the *no group monotonicity* requirement, without imposing *group anti-monotonicity*. But we think that our notion of joint action is adequate for the applications we have in mind, related to the specification of organisations and of the acts and joint direct acts of the agents involved, where we want to express, for instance, that two agents make a contract, or cases where the signature of two administrators of a company is necessary in order to produce some specified result.

3.2 Collective agents and their actions

We have seen how to represent a notion of collective agency associated to a joint act of a group of agents, where each member of the group has acted and the actions of all the members of the group were necessary to produce some result.

But there is another kind of agency (with particular interest to us) that is associated with groups of agents, and usually also called of *collective agency*, that has to do with the agency of groups where we may say that the group as a whole has acted (and produced some result, like making a contract), without meaning that all members of the group have acted, and this may happen because the group has accepted rules according to which, under certain circumstances, someone can act in the name of the group. The group acts through the direct act of some member (or of someone with a specific relationship with the group), or through the joint direct act of some of its members, that may be said to act on its behalf, according to the rules that are accepted.

We are not thinking of a group of agents that act sporadically, but on a group of agents that have some common goals and have some more or less permanent activity. In general, the specific members of the group may even change, without changing such activity, and so the best way to see things is as if the group has created a new entity that represents the group, but that is not equal to the set of the members of the group (that can change, without changing the nature of such entity). Such entity can be informal, or more formal, even with legal recognition, as it is the case of the organisations. Although we are particularly interested in the latter cases, we think that the main general ideas below also apply (with some adjustments) to the other more informal cases.

An organisation should be regarded as different from the set of agents that are related to it. We may see an organisation (as other institutionalized collective entities) as an agent (that we may call a *collective agent)*, that have a proper identity (for instance, the set of workers in a company can change, but the company will remain the same), that can act and interact with the external world (with the other agents), making for instance contracts, and that can be subject of obligations and be (even legally) responsible by its violation / non-fulfilment. However, a collective agent is different from the human agents by the fundamental fact that it cannot act, directly: someone has to act in its name. (The organisation must act through the acts of other agents.)

But how can we represent that some agent is acting in the name of an organisation? As stated in [26], within institutions, organisations, 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). And, to express such ''*count as'' (''meaning'') relations*, Jones and Sergot proposed in [26] the use of a conditional modal operator[38] $\Rightarrow_s$ (where the index s refers the relevant system of analysis). They also proposed a normal modal operator, herein denoted by R_s, where expressions of the form $\mathrm{R}_s\varphi$ may be read as follows: "according to the rules/norms operating/accepted in institution s, φ is the case" or "it is *recognised*/accepted by institution s that φ is the case". Using these operators, we could express, for instance, that "a's act of bringing about φ counts, within the institution s (where s may be the organisation o or the competent legal system) as a means by which the organisation o establishes state of affairs ψ"[39], as follows

$$\mathrm{E}_a\varphi \Rightarrow_s \mathrm{E}_o\psi \text{ (or by } \mathrm{E}_a\varphi \rightarrow \mathrm{R}_s\mathrm{E}_o\psi)$$

(where, as we have stated at the introduction, we use E_a to generically refer to an operator meaning "agent a 'sees to it' that", independently of its specific definition).

However, although an organisation may establish that some specific agent a may act on behalf of the organisation, for bringing about some specific state of affairs, this is not practical and it is not the general procedure used by organisations. In order to make the things work, particular relationships are established between the organisation and other agents, by which the latter inherit (some of) the obligations of the former, and acts on its name. Some of those relationships are stable, and intended to be part of the structure of the organisation, corresponding to what is usually called of *posts*, or *roles* within the organisation, and the statute of the organisation distributes the duties of the organisation 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), and describing who has the power to act in the name of the organisation. On the other hand, those that can act in the name of an organisation can establish new obligations for the organisation through their acts, for instance by establishing contracts with other agents (persons, organisations, etc.). And in this way we have a dynamic of obligations, where the obligations flow from the organisation to the holders of some roles, and these, through their acts, create new obligations in the organisation.

But an agent can be the holder of different roles within the same organisation, or in different organisations, and can *act by playing different roles*. Moreover, 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 the company car when on duty – i.e. when he is acting in the position 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 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)[40]. Thus it becomes essential to have a way of expressing the role an agent played when he/she acted.

For this reason, in [13, 31] we have extended the "sees to it" action operators, expressing the role r that the agent a has played when he, or she, has brought it about that φ, through

[38]Such operator is further studied, developed and applied by Davide Grossi in [20].

[39]Assertions φ and ψ may be the same (e.g."sell organisation's car"), or not (e.g. ψ might represent "sell organisation's car" and φ the means, like "sign an appropriate document").

[40]As another example, we note that a same person can be the chief editor of two journals, and he or she might reject a paper as chief editor of one journal, while at the same time he, or she, might accept the same paper as chief editor of the other journal.

sentences of the form[41] $E_{a:r}\varphi$. We have introduced, in the proposed formal language, a way to construct the role terms, we have characterised the properties of such extended action operator, and, following the approach of Kanger, Pörn and Lindahl, we have combined such operators with deontic operators, in order to specify organisations and agents' interaction in general.

But it is not enough to have a way to express the role an agent has played when he has acted. It is fundamental to know which acts *count as* acts in a particular role. And, on the other hand, *in reality what we have is agents directly acting*. Thus, the precise question is: in what conditions such direct acts will be recognised as acts in some role by the organisation, society, normative system, etc.?

In summary, agents (and groups of agents) can act directly, but what we have called of collective agents (like organisations) no; they always act indirectly through the acts of other agents. This may happen because, e.g. by a contract, the collective agent gives powers to a specific agent to do a specific (direct) act in its name, or because the statute of the collective agent attributes such power to some role (like the role of president), meaning that when some agent does a direct act that is recognised as an act playing that role that will count as an act of the collective agent.

These considerations are only an illustration (within the applications we have in mind) of the interest of having means of stating the direct and immediate effects of the agents' own actions, allowing us to discriminate such direct and immediate effects from those indirect effects that follow from them, either sometime later, by some causal connection, or[42] (we may suppose immediately) by institutional connection (by the norms applicable e.g. within the relevant organisation). And we may try to conceive a "direct" action operator (that we can generically denote by D) to express such direct effects (and continue to use the operator E to express states of affairs that are brought about by the agents, in some generic sense).

4 The direct effects of agents' actions

Let us then see how to represent the direct, in the sense of immediate, effects of the agents' actions, within the stit framework. In what follows we will consider the simpler case of one agent, but we think that the ideas can be adapted also to sets of agents (like in section 3.1).

Consider first the deliberative stit, and suppose that we want to express that an agent a has made (at the present moment m) a choice $C_{m,a}(h)$ that has φ (true) as an immediate effect, which we will denote by $D_a^{\rightarrow}\varphi$. By this we do not intend to mean that φ is the case at m (according to h). What we mean is that immediately after m, φ must be the case.

One possible way of expressing this is as follows:

- $\mathsf{M} \models_{h,m} D_a^{\rightarrow}\varphi$ iff
  (1) for every $h' \in C_{m,a}(h)$, there is $m^* \in h'$ such that $m^* > m$ and, for any $m < m' \leq m^*$, $\mathsf{M} \models_{h',m'} \varphi$
  (2) (and) there exists $h' \in H_m$ such that it is not the case that there is $m^* \in h'$ such that $m^* > m$ and, for any $m < m' \leq m^*$, $\mathsf{M} \models_{h',m'} \varphi$

[41] In [13, 31], and herein (in what follows), the operator $E_{a:r}$ is considered in the achievement sense, i.e. concerning the past actions/choices of the agent a.

[42] The case with interest to us within our applications.

If we equip (extend) our language with linear temporal operators (like the until operator U^{ee}, already defined), then it becomes much simpler to present the definition above, being possible to define the direct action operator through the following abbreviation

$$D_a^{\rightarrow}\varphi =_{\text{def}} [a\ dstit: \varphi U^{ee}\varphi]$$

and this was precisely the approach followed by Mark Brown in [10], where he proposes a *pstit* operator (with *p* for *progressive tense*) defined as $[a\ dstit{:}\ W\varphi]$. (In [10] the until operator was not considered, and so $W\varphi$ was not defined as $\varphi U^{ee}\varphi$, W being considered a primitive linear operator.)

Suppose now that we want to express that an agent has *just brought* it about that φ, which we will denote by $D_a^{\leftarrow}\varphi$, thus concerning the immediate past actions of the agent. This suggests that we turn our attention to the achievement stit, instead of the deliberative stit, but a similar strategy of combining (now) the *astit* operator with a tense operator does not seem obvious.

The Chellas's operator Δ_a resembles what we want. But there is a problem, since Chellas's semantics for Δ_a does not provide the moment where the relevant action took place, and where the agent could act differently (allowing us to add a counterfactual condition, that we think essential). One idea might be to implement in the *astit* semantics a *nearly now* requirement, as it was suggested in [16], in order to solve a problem with the stit-past sentences[43]. The idea is to place witnesses as close as possible to the moment of agency. Such a requirement would mean that [16, pp 514] ‘‘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’’.

One option to try to implement such requirement might be to add to the *astit* semantic definition

- $\mathsf{M} \models_{h,m} D_a^{\leftarrow}\varphi$ iff there is a moment $m^* < m$ such that
  (1) $\mathsf{M} \models_{h',m@h'} \varphi$ for every history $h' \in C_{m^*,a}(h)$
  (2) $\mathsf{M} \not\models_{h',m@h'} \varphi$ for some history $h' \in H_{m^*}$

one of the following conditions (each one implied by the next one)

(3-i) $\forall_{m^*<m'<m}\forall_{h'}(h' \in H_{m'} \rightarrow h' \in C_{m',a}(h))$
(3-ii) $\forall_{m^*<m'<m}\forall_{h^*}\forall_{h'}(h^* \in C_{m^*,a}(h) \wedge h' \in H_{m'@h^*} \rightarrow h' \in C_{m'@h^*,a}(h^*))$
(3-iii) $\forall_{h^*}(h^* \in C_{m^*,a}(h) \rightarrow \forall_{h'}\ (\exists_{m^*<m'\leq m} h' \in H_{m'@h^*} \rightarrow h' \in H_{m@h^*}))$
(3-iv) $\forall_{h^*}\forall_{h'}(h^* \in H_{m^*} \wedge \exists_{m^*<m'\leq m} h' \in H_{m'@h^*} \rightarrow h' \in H_{m@h^*})$

Informally: according to (3-i), if a follows h, then all choices of a between (the choice point) m^* and m are vacuous choices (i.e. a does not really act between m^* and m); according to (3-ii), if a takes the choice $C_{m^*,a}(h)$, at m^*, then a does not really act between m^* and the time instant of m; according to (3-iii), if a takes the choice $C_{m^*,a}(h)$, at m^*, then no history within that choice split between m^* and m (it is like if no agent really act between m^* and the time instant of m); and, according to (3-iv), no agent really act between moment m^* and the time instant of m.

However, we are not going to proceed this line of research in this paper. The best way of capturing such *nearly now* requirement constitutes matter that still needs further work.

[43]Problem described in [16, pp 511] as follows: ‘‘in stit theory it is possible for an agent to see to something's being so in the past without ever in the past seeing to it that it is so’’.

Here, we just want to note that if we are concerned with applications where the time is discrete (where we may assume that the set of relevant, or observable moments is discrete), and we consider the time to be isomorphic to the set of integers, as in [14] (or isomorphic to the set of natural numbers, as in [6, 7, 43]), then the things become much simpler, it being possible to treat similarly both operators $D_a^{\rightarrow}$ and $D_a^{\leftarrow}$, as follows (where we write $m-1$ to refer to the moment before m, moment that is equal to $m-1@h$ if history h pass through m, and $m+1@h$ to refer to the moment after m in history h):

- $\mathsf{M} \models_{h,m} D_a^{\leftarrow} \varphi$ iff[44] (it exists $m-1$, in the natural time's case, and)
  (1) $\mathsf{M} \models_{h',m@h'} \varphi$ for every history $h' \in C_{m-1,a}(h)$, and
  (2) $\mathsf{M} \not\models_{h',m@h'} \varphi$ for some history $h' \in \mathrm{H}_{m-1}$
- $\mathsf{M} \models_{h,m} D_a^{\rightarrow} \varphi$ iff
  (1) $\mathsf{M} \models_{h',m+1@h'} \varphi$ for every history $h' \in C_{m,a}(h)$, and
  (2) $\mathsf{M} \not\models_{h',m+1@h'} \varphi$ for some history $h' \in \mathrm{H}_m$

Within this discrete time, we can also define both these operators as simple iterations of *dstit* and temporal linear operators (the previous time and next time operators defined in section 2.2) as follows

- $D_a^{\leftarrow} \varphi =_{\text{def}} Y[a \textit{ dstit: } X\varphi]$
- $D_a^{\rightarrow} \varphi =_{\text{def}} [a \textit{ dstit: } X\varphi]$

Related to this topic, we should also mention a very recent work of Jan Broersen [8], where a stit logic, with a new semantics, is proposed, whose distinguished feature is that "actions only take effect in 'next' states, where 'next' refers to immediate successors of the present state", and where one of the operators proposed ([*a xstit*]) has some similarities with $D_a^{\rightarrow}$. There are however important differences between the approaches here and in [8], and we leave a comparison between them for future work.

5 Dynamics

The "sees to it" operators are essentially static operators, that do not allow us to relate what happens in the moment/state the agent initiates an action and what happens in the resulting state of the action. Although this has virtues, it also has some limitations. In particular, we would like to have an easy way to express that if some direct action of seeing to it type is performed, something obtains (allowing us to make hypothetical reasoning), similarly to what it is provided by the dynamic logic operators.

Suppose, for the moment, that the time is discrete, and suppose that we write $[D_a^{\rightarrow}\varphi]\psi$ 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 $D_a^{\rightarrow}\varphi \rightarrow D_a^{\rightarrow}\psi$ does not express what we want. When we write $[D_a^{\rightarrow}\varphi]\psi$ we do not intend to state that the agent a also "sees to it" that ψ. In particular, we want that our operator $[D_a^{\rightarrow}\varphi]$ will be a normal modal operator. Moreover, we want to state that ψ would also obtain, if the agent a has selected any other action (different from the expressed by the choice $C_{m,a}(h)$) that would also be an action of seeing to it that φ.

[44]Note that if we just consider (1), and define $R_{m,a}(h,h')$ as $h' \in C_{m-1,a}(h)$, we get Chellas's "sees to it" operator. Note also that the stit-past sentences problem, referred in [16], does not exist with this operator, since (where P means "sometime in the past") we have $\models \neg D_a^{\leftarrow} P\varphi$.

The definition we gave in [12] was

- $[D_a^{\rightarrow}\varphi]\psi =_{def} \Box([a\ cstit\text{: } D_a^{\rightarrow}\varphi] \rightarrow [a\ cstit\text{: } X\psi])$

that corresponds to the following semantic definition (without considering temporal operators)

- $\mathsf{M} \models_{h,m} [D_a^{\rightarrow}\varphi]\psi$ iff for every $h' \in H_m$:
  if $\mathsf{M} \models_{h'',m} D_a^{\rightarrow}\varphi$ for every history $h'' \in C_{m,a}(h')$
  (i.e. if the choice $C_{m,a}(h')$ corresponds to an action of directly seeing to it that φ),
  then $\mathsf{M} \models_{h'',m+1@h''} \psi$ for every history $h'' \in C_{m,a}(h')$

However, the previous definition can be simplified. As a matter of fact, since

$$\models [a\ cstit\text{: } D_a^{\rightarrow}\varphi] \leftrightarrow D_a^{\rightarrow}\varphi$$

and since $\models \Box([a\ cstit\text{: } \varphi_1] \rightarrow [a\ cstit\text{: } \varphi_2]) \leftrightarrow \Box([a\ cstit\text{: } \varphi_1] \rightarrow \varphi_2)$, for any formulae φ_1 and φ_2 (see appendix), we can simply define

- $[D_a^{\rightarrow}\varphi]\psi =_{def} \Box(D_a^{\rightarrow}\varphi \rightarrow X\psi)$

that corresponds to the following much simpler semantic definition

- $\mathsf{M} \models_{h,m} [D_a^{\rightarrow}\varphi]\psi$ iff, for every $h' \in H_m$, if $\mathsf{M} \models_{h',m} D_a^{\rightarrow}\varphi$, then $\mathsf{M} \models_{h',m+1@h'} \psi$

As it would be expectable, the two direct action operators are related by the formula

$$[D_a^{\rightarrow}\varphi]D_a^{\leftarrow}\varphi$$

Before proceeding, we can very briefly compare this kind of dynamic operator with the delta-operator of Krister Segerberg. In [35], Segerberg proposed an operator δ that, applied to a sentence φ, provides an action term representing an action read ''the bringing about of φ'' or ''doing φ''. He used dynamic logic for its characterisation, proposing the axioms $[\delta(\varphi)]\varphi$ and $[\delta(\varphi)]\phi \rightarrow ([\delta(\psi)]\varphi \rightarrow [\delta(\psi)]\phi)$. Although we also have $[D_a^{\rightarrow}\varphi]\varphi$, we do not have (as valid) the schema $[D_a^{\rightarrow}\varphi]\phi \rightarrow ([D_a^{\rightarrow}\psi]\varphi \rightarrow [D_a^{\rightarrow}\psi]\phi)$, since, for instance, we might have a situation where $[D_a^{\rightarrow}\varphi]\bot \wedge \neg[D_a^{\rightarrow}\psi]\bot \wedge [D_a^{\rightarrow}\psi]\varphi$ is true at (h,m) (which implies that $[D_a^{\rightarrow}\varphi]\phi \wedge [D_a^{\rightarrow}\psi]\varphi$ is also true at (h,m)), and from it we cannot deduce the truth, at (h,m), of $[D_a^{\rightarrow}\psi]\phi$.

Although we are here mainly interested in applying these kind of dynamic operator [_] to these "(direct) seeing to it" action sentences, we can generalise this operator, allowing it to apply to any sentence (and labelling it with an agent), in formulae of the form $[a\text{: }\varphi]\psi$, 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''). Formally[45]:

- $[a : \varphi]\psi =_{def} \Box([a\ cstit\text{: } \varphi] \rightarrow X\psi)$

or directly, semantically (without considering temporal operators in our language)

[45]The definition we gave in [12] was $[a\text{: } \varphi]\psi =_{def} \Box([a\ cstit\text{: } \varphi] \rightarrow [a\ cstit\text{: } X\psi])$. But, as before, taking into account that $\models \Box([a\ cstit\text{:}\varphi_1] \rightarrow [a\ cstit\text{: } \varphi_2]) \leftrightarrow \Box([a\ cstit\text{: } \varphi_1] \rightarrow \varphi_2)$, for any formulae φ_1 and φ_2, such definition can be simplified as it is referred.

- $\mathsf{M} \models_{h,m} [a\colon \varphi]\psi$ iff
  for every $h' \in \mathrm{H}_m$, if $\mathsf{M} \models_{h'',m} \varphi$ for every h'' such that $h'' \in C_{m,a}(h')$, then $\mathsf{M} \models_{h',m+1@h'} \psi$

And, using this operator, we can define $[\mathrm{D}_a^{\rightarrow}\varphi]\psi$ as an abbreviation of $[a\colon \mathrm{D}_a^{\rightarrow}\varphi]\ \psi$.

With respect to the logic of this kind of dynamic operator $[a\colon \varphi]\psi$, we have that it is a normal logic with respect to ψ, and, with respect to φ, it verifies the RE (replacement of equivalents) rule, but not the RM-rule, and it satisfies, among others, the following *augmentation principle*:

$$[a:\varphi]\psi \rightarrow [a:\varphi\wedge\varphi']\psi \quad \text{(note that if } \models \varphi\rightarrow\phi \text{ then } \models [a:\phi]\psi\rightarrow[a:\varphi]\psi)$$

On the other hand, the logic does *not* satisfy any of the following schemas (i.e., they are not valid schemas):

$[a:\varphi]\psi \rightarrow \psi$
$[a:\varphi]\psi \rightarrow \varphi$ (neither satisfies $[a:\varphi]\psi \rightarrow \Diamond\varphi$)
$[a:\varphi]\psi \rightarrow <a:\varphi>\psi$ (where $<a:\varphi> = \neg[a:\varphi]\neg$, i.e. $<a:\varphi>$ is the dual of $[a:\varphi]$)

The non-validity of the schema $[a\colon \varphi]\psi \rightarrow \psi$ is crucial in order that this operator can be used for agent decision. For instance, in a moment where

$$\text{"}a \text{ is not liable to punishment"} \wedge [\mathrm{D}_a^{\rightarrow}\varphi]\ \text{"}a \text{ is liable to punishment"}$$

is the case, the agent a should not take the decision of bringing about φ.

Note also that we must be careful in interpreting[46] $<a\colon \varphi>\psi$, i.e. the dual of $[a\colon \varphi]\psi$. The formula $<a{:}\varphi>\top$ means that it is possible for agent a to choose φ. However, the formula $<a\colon \varphi>\psi$ 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, that we can denote by $<[a\colon \varphi]>\psi$, defined as follows:

- $<[a\colon \varphi]>\psi =_{\mathrm{def}} \Diamond([a\ \mathit{cstit}\colon \varphi] \wedge [a\ \mathit{cstit}\colon X\psi])$

Naturally $[a:\varphi]\psi \wedge <a:\varphi>\top$ implies $<[a:\varphi]>\psi$.

Finally, note that we may define similar operators without imposing that time must be discrete[47]. We just have to replace X by W in the abbreviations above:

- $[a:\varphi]\psi =_{\mathrm{def}} \Box([a\ \mathit{cstit}\colon \varphi] \rightarrow W\psi)$ (that is equivalent to $\Box([a\ \mathit{cstit}\colon \varphi]\rightarrow[a\ \mathit{cstit}\colon W\psi])$)
- $<[a\colon \varphi]>\psi =_{\mathrm{def}} \Diamond([a\ \mathit{cstit}\colon \varphi] \wedge [a\ \mathit{cstit}\colon W\psi])$

We end, by giving some very brief illustrations of the use of these dynamic operators to express some relevant facts within the applications we have in mind (related to organisation's specification).

[46] $<a:\varphi>\psi = \neg[a:\varphi]\neg\psi$, formula that it is equivalent to $\Diamond([a\ \mathit{cstit}\colon \varphi] \wedge\neg X\neg\psi)$, which is equivalent (if we assume that no history has a last moment) to $\Diamond([a\ \mathit{cstit}\colon \varphi] \wedge X\psi)$.

[47] We should stress, however, that if we allow that there might exist histories with a last moment, then $[a\colon \varphi]\psi$ is no longer a normal operator with respect to ψ, since in such case $[a\colon \varphi]\top$ is not valid (for formulae φ not imposing the existence of later moments).

In order for someone act, playing some role, he or she must be qualified to play that role. This was a basic principle of our approach and in [13, 31] we have expressed it by the following schema

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

where qual(a:r) is true if and only if the agent a holds the role r – "agent x is qualified to play the role r"[48]. As particular instances of this principle, we have e.g. $E_{a:president-of(o)}\varphi \rightarrow$ *is-president_ of(a,o)* that means that in order to a bring about that φ is the case, playing the role of president of the organisation o, agent a must be the president of the organisation o.

In most cases the proposed characterisation of such principle works well, because the qualifications of the agent does not change by his (referred) act. But suppose now that an agent a playing the role of owner of building *xpto*, sells that building, which we can express by $E_{a:owner-of(xpto)}$ *sells(xpto)*. According to (*), we have then that

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

but this makes no sense. After a has sold the building, a is no longer the owner of that building.

In cases like this, it seems that the real relevant moment to evaluate the qualification is immediately before the act. And we can try to express our general requirement that *in order to an agent act playing some role, he must be qualified to play that role*, by making use of our dynamic operators and of the (already referred) recognition operator R_s, for instance through the sentence (instead of (*)):

$$< [a : \top] > R_s E_{a:r}\varphi \rightarrow \text{qual}(a : r)$$

that means that if the agent a can make an action that will be recognised by the society (or the normative system) s as an act in the role r (for bringing about some state of affairs), then a must be qualified to play that role.

On the other hand, to know that someone is acting at playing some role, it might be not enough to know that he is qualified to play that role, because he may be qualified to play many roles. Thus, we have the question of knowing in what conditions some direct act will be recognised as an act playing some role by an organisation, the society, the normative system, etc.? In general, there are some conventional signals that an agent exhibit that guarantees (or that the agent must exhibit in order to guarantee) that the act that he, or she, will perform will be recognised as an act playing some role. And we can try to use the previous operators also to represent such situations.

Consider again the example of the owner of a building xpto. Probably we can state that

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

(where *doc-own(a,xpto)* means "a has exhibited the document of ownership of building xpto"), meaning that when 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 will be recognised (or assumed) as an act playing the role of owner of that building.

[48] In general, given a role $r(\ldots)$, we have that $qual(a,r(\ldots))$ is a predicate that can be described as $is\text{-}r(a,\ldots)$. For instance, *qual(a,president_ of(o)) = is-president_ of(a,o)*. See [13, 31] for details.

In this example, probably we could even write that

$$<[a:\top]> \mathrm{R}_s\mathrm{E}_{a:owner\text{-}of(xpto)}\varphi \rightarrow \mathrm{D}_a^{\leftarrow}\, doc\text{-}own(a, xpto)$$

a sentence that would mean that if it is possible for a to make an action that will be recognised 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[49].

6 Sergot's recent approach to "sees to it" / "brings it about"

We conclude with a brief reference to a different approach to the ''sees to it'' / ''brings it about'' concept, recently proposed by Marek Sergot.

In [38], Sergot presents a framework for describing and analysing norm-governed multiagent systems, where it is distinguished between system norms and agent-specific norms, and where it is possible to identify and characterise several different categories of non-compliant behaviour. The formal (modal-logical) language proposed includes operators for expressing that a particular agent brings it about that such-and-such is the case[50], getting a logic that is said to bear a strong resemblance to Ingmar Pörn's [33] logic of ''brings it about'' action, except that there is a switch from talking about an agent's bringing about a certain state of affairs to an agent's bringing about that a transition has a certain property. In what follows we will restrict our attention to this novel feature, starting by briefly presenting the underlying formal framework.

The semantics is based on what is called an *agent-stranded labelled transition system* (*agent-stranted LTS*, for short) that is a structure of the form

$$(\mathrm{S}, \mathrm{A}, \mathrm{R}, prev, post, label, \mathrm{Ag}, strand)$$

where: S and R are disjoint, non-empty sets of *states* and *transitions* respectively; A is a set of *transition labels* (also called *events*); *prev* and *post* are functions from R to S ($prev(\tau)$ denotes the initial state of a transition τ and $post(\tau)$ its resulting state); *label* is a function from R to A; Ag is a (finite) set of agent names; and *strand* is a function on Ag x A. The idea of the latter (new) component is that $strand(a,\varepsilon)$ picks out from a transition label/event ε the component or 'strand' that corresponds to agent a's contribution to the event ε. We can write ε_a for $strand(a,\varepsilon)$ (ε_a may represent several concurrent actions by a, or actions with non-deterministic effects). We can also speak of a's strand of a transition τ, which is denoted by τ_a and defined as $strand(a,label(\tau))$. It is referred that τ_a may be thought of as the actions of agent a in the transition τ, where this does *not* imply that τ_a necessarily represents deliberate action, or action which has been freely chosen by a.

[49] Like, when acting as an administrator of a computer system, a user must authenticate by inserting the respective password.

[50] Sergot uses the expression 'brings it about', and not 'sees to it', because (following Hilpinen) he considers that the expression 'seeing to it' that φ usually characterises deliberate, intentional action, and 'bringing it about' that φ does not have such a 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, and the concept of agency he tries to capture is of the latter ('brings it about') kind. Herein, following Mark Brown (see [10, footnote 1]), we are not considering such distinction between the expressions 'brings it about' and 'sees to it'.

A two-sorted language is employed, with *state formulae* and *transition formulae*. There is a set of propositional *state atoms*[51], for expressing properties of states, and a disjoint set of propositional atoms for expressing properties of events and transitions, called *event atoms*. The state and transition (propositional) formulae are built from the correspondent atoms through the usual truth-functional connectives. This two-sorted propositional logic is then extended (as described below) with some modal operators that transform formulae of one sort in formulae of the other sort. We will use φ_S (ψ_S, etc.) when we want to stress that we are generically referring to a state formula φ, and φ_T to generically denote a transition formula φ. However, we can write simply φ when it is clear from the context what is the sort of formula φ (for instance, from the modal operator we are applying to it)[52].

The models M extend the *agent-stranted LTS* structures with two *valuation* functions, one, h^f, that maps each state atom into a subset of S, and another,h^a, that maps each event atom into a subset of A. The truth-value of the state formulae is evaluated with respect to a state s and the truth-value of the transition formulae is evaluated with respect to a transition τ[53], and a transition τ is said to be of type φ (for φ a transition formula) if $\mathsf{M} \models_\tau \varphi$.

As referred to above, the base two-sorted propositional logic is extended with modal operators. Concretely, the language is first extended as follows: $[\varphi_T]\psi_S$ is a state formula that expresses that the state formula ψ_S is satisfied in every state following a transition of type φ; and $0{:}\varphi_S$ and $1{:}\varphi_S$ are transition formulae that are satisfied by a transition τ when its initial state satisfies φ_S and its resulting state satisfies φ_S, respectively. Formally:

$\mathsf{M} \models_s [\varphi_T]\psi_\mathrm{S}$ iff[54] $\mathsf{M} \models_{post(\tau)} \psi_S$ for every transition τ such that $prev(\tau) = s$ and $\mathsf{M} \models_\tau \varphi_T$
$\mathsf{M} \models_\tau 0{:}\varphi_S$ iff $\mathsf{M} \models_{prev(\tau)} \varphi_\mathrm{S}$
$\mathsf{M} \models_\tau 1{:}\varphi_S$ iff $\mathsf{M} \models_{post(\tau)} \varphi_\mathrm{S}$

(According to what we said above, we can simply write $[\varphi]\psi$, $0{:}\varphi$ and $1{:}\varphi$, instead of $[\varphi_T]\psi_S$, and $0{:}\varphi_S$ and $1{:}\varphi_S$.)

And, in order to introduce the agency operators (''brings it about''), the language is also extended with the following modal operators (for every a in Ag)[55]:

$\mathsf{M} \models_\tau [\mathrm{alt}]\varphi_\mathrm{T}$ iff $\mathsf{M} \models_{\tau'} \varphi_\mathrm{T}$ for every (transition) τ' such that $prev(\tau') = prev(\tau)$
$\mathsf{M} \models_\tau [a]\varphi_\mathrm{T}$ iff $\mathsf{M} \models_{\tau'} \varphi_\mathrm{T}$ for every τ' such that $prev(\tau') = prev(\tau)$ and $\tau'_a = \tau_a$
$\mathsf{M} \models_\tau [\backslash a]\varphi_\mathrm{T}$ iff $\mathsf{M} \models_{\tau'} \varphi_\mathrm{T}$ for every τ' such that $prev(\tau') = prev(\tau)$ and $\tau'_b = \tau_b$ for every $b \in \mathrm{Ag}\text{-}\{a\}$

Finally, the following agency operators are introduced as abbreviations, as follows:

$$\mathrm{E}_a\varphi_\mathrm{T} =_{\mathrm{def}} [a]\varphi_\mathrm{T} \wedge \neg[alt]\varphi_\mathrm{T}$$
$$\mathrm{E}_a^+\varphi_\mathrm{T} =_{\mathrm{def}} [a]\varphi_\mathrm{T} \wedge \neg[\backslash a]\varphi_\mathrm{T}$$

[51] As in [39] we will not discriminate here the structure of the atoms.

[52] In [38, 39] it is used F to generically denote a state formula and φ to generically denote a transition formula. However, we prefer to not follow such convention, by uniformity with the rest of this paper, where we use φ both to denote formulae that within this approach would be consider as state formulae, and formulae that would be probably considered as transition formulae. Also by uniformity with the rest of this paper, we use a and b to refer to agents, and not x and y as in [38, 39].

[53] If p is a state atom, then $\mathsf{M} \models_s p$ iff $s \in h^f(p)$; if α is an event atom, then $\mathsf{M} \models_\tau \alpha$ iff $label(\tau) \in h^a(\alpha)$.

[54] There is some typo in the formal definition given for $\mathsf{M} \models_s [\varphi_T]\psi_S$ in [38, 39]. But we think that the definition we are giving expresses what it is wanted by Sergot.

[55] It is said in [38] that the language can be generalized to allow expressions of the form $[\mathrm{G}]\varphi$, for any $\mathrm{G} \subseteq \mathrm{Ag}$, topic that is then analysed in [39], where various possible group agency operators are discussed.

where, informally, $E_a\varphi_T$ means that agent a brings it about that φ, and $E_a^+\varphi_T$ expresses that it is a, and *a alone*, who brings it about that φ (where φ is a transition formula).

Note that we have (as valid) the schemas $E_a^+\varphi \to E_a\varphi$ and, if $b \neq a$, $E_a^+\varphi \to \neg E_b\varphi$.

Having introduced the main features of Sergot's framework and logic, we can now make a brief comparison with stit framework and our proposals here.

The first obvious difference, regarding the agency operators, is that $E_a\varphi$ (or $E_a^+\varphi$) represents that agent a brings it about that a transition is of type φ, and usually in the logics of agency what it is stated is that an agent brings it about (or "sees to it") a certain state of affairs (thus a state formula φ_S, and not a transition formula φ_T). But that can be expressed within the proposed framework as $E_a 1{:}\varphi_S$.

As it is said in [39, section 4.4], a formula like $E_a 1{:}\varphi_S$ "expresses *one sense* in which it might be said that a 'brings about' such-and-such a state of affairs φ exists. It is not the only sense, because it says that φ holds in the state immediately following the transition, whereas we might want to say merely that φ holds at some (unspecified) state in the future. Logics of agency usually do not insist that what is brought about is immediate; indeed, since transitions are not elements of the semantics, references to 'immediate' or the 'next state' are not meaningful"[56]. (Thus, as we discuss below, $E_a 1{:}\varphi$ may be seen to correspond more to ours $D_a^{\rightarrow}\varphi$).

Comparing stit and Sergot semantic frameworks, and simplifying some things, Sergot's states may be seen as corresponding to the moments, and a transition τ starting at a state/moment m (i.e. such that $prev(\tau) = m$) may be seen as corresponding to an history h passing though the moment m (*plus* all histories that are undivided from h at that moment m). Following this idea, and assuming that time in stit is discrete (isomorphic to the set of integers, or to the set of natural numbers), we can try to "(re)define" Sergot's proposal within stit framework as follows (where the truth value of any formula is evaluated with respect to a pair (h,m), such that h passes through m, "identifying" a transition τ with a pair (h,m), such that m=$prev(\tau)$, and "identifying" τ_a with $C_{m,a}(h)$)[57]:

i) $\mathsf{M} \models_{h,m} [\varphi]\psi$ iff $\mathsf{M} \models_{h',m+1@h'} \psi$ for every $h' \in H_m$ such that $\mathsf{M} \models_{h',m} \varphi$
ii) $\mathsf{M} \models_{h,m} 0{:}\varphi$ iff $\mathsf{M} \models_{h,m} \varphi$
iii) $\mathsf{M} \models_{h,m} 1{:}\varphi$ iff $\mathsf{M} \models_{h,m+1@h} \varphi$
iv) $\mathsf{M} \models_{h,m} [\text{alt}]\varphi$ iff $\mathsf{M} \models_{h',m} \varphi$ for every $h' \in H_m$
v) $\mathsf{M} \models_{h,m} [a]\varphi$ iff $\mathsf{M} \models_{h',m} \varphi$ for every $h' \in C_{m,a}(h)$
vi) $\mathsf{M} \models_{h,m} [\backslash a]\varphi$ iff $\mathsf{M} \models_{h',m} \varphi$ for every h' such that $h' \in \bigcap_{b \in Ag-\{a\}} C_{m,b}(h)$

According to this approximation to Sergot's proposal (within stit framework), we have that:

i) $[\varphi]\psi$ is equivalent to $\Box(\varphi \to X\psi)$
ii) $0{:}\varphi$ is equivalent simply to φ
iii) $1{:}\varphi$ is equivalent to $X\varphi$
iv) $[\text{alt}]\varphi$ is equivalent to $\Box\varphi$
v) $[a]\varphi$ is equivalent to [a *cstit:* φ]

[56] Another essential difference, according to [39], is that $E_a 1{:}\varphi$ (and $E_a^+ 1{:}\varphi$) are *transition* formulae, and "cannot be used to say that in a particular state s, a brings it about that such-and-such a state of affairs φ holds". Sergot discuss some possibilities on how to express that, but he does not develop such ideas.

[57] If p is a state atom, then $\mathrm{M} \models_{h,m} p$ iff $m \in h^f(p)$; if α is an event atom, then $\mathsf{M} \models_{h,m} p$ iff *label*$(<h,m>) \in h^a(\alpha)$.

vi) $E_a\varphi$ is equivalent to [*a dstit:* φ]
vii) and $E_a 1{:}\varphi$ is equivalent to ours $D_a^{\rightarrow}\ \varphi$.

Note also that if we assume the *something happens condition*, then the schema $E_a\varphi \rightarrow \Diamond E_a^{+}\varphi$ becomes valid.

(Proof: Suppose that (i) $\mathsf{M} \models_{h,m} E_a\varphi$ and let $h' \in H_m$ be such that $\mathsf{M} \not\models_{h',m} \varphi$, whose existence follows from $\mathsf{M} \models_{h,m} \neg[\text{alt}]\varphi$. Let $X = (\cap_{b\in Ag-\{a\}} C_{m,b}(h')) \cap C_{m,a}(h)$.

Assuming the something happens condition, $X \neq \varnothing$. Let $h^* \in X$.

Then $h^* \in C_{m,a}(h)$, and so $C_{m,a}(h^*) = C_{m,a}(h)$. Thus (since (i) implies that $\mathsf{M} \models_{h,m} [a]\varphi$):

$$\mathsf{M} \models_{h'',m} \varphi \text{ for every } h'' \in C_{m,a}(h^*), \text{ i.e. (iii) } \mathsf{M} \models_{h^*,m} [a]\varphi$$

Since $h^* \in X$, we also have that $\forall_{b\in Ag-\{a\}}\ C_{m,b}(h^*) = C_{m,b}(h')$. Thus:

$$h' \in H_m \text{ is such that } M \not\models_{h',m} \varphi \text{ and } h' \in \cap_{b\in Ag-\{a\}} C_{m,b}(h^*), \text{ i.e. (iv) } \mathsf{M} \models_{h^*,m} \neg[\backslash a]\varphi$$

But then (by (iii) and (iv)), we have that $\mathsf{M} \models_{h^*,m} E_a^{+}\varphi$, which implies (since both h and h^* pass through m) that $\mathsf{M} \models_{h,m} \Diamond E_a^{+}\varphi$, as we wish to prove.)

Naturally, in this kind of ''translation'' (or (re)interpretation) of Sergot's proposal, we lose part of Sergot's ideas and motivation (and his distinction between state and transition formulae). For instance, in Sergot's approach, in [φ]ψ the formula φ must be a transition formula. Here we have not considered transition atoms; within our approach, we may see (informally) the ''transition assertions'' as being either temporal assertions, related to the next time operator (assuming that we consider such operator and discrete time), or assertions about the choices/actions of the agents. And the dynamic operators herein proposed are precisely of such kind. For instance, our [*a*: φ]ψ could be seen as Sergot's [[*a*]φ]ψ, and [*a*]φ is a transition formula[58], and <[*a*: φ]> ψ can be formulated as <alt>([*a*]φ∧[*a*]1:ψ).

We are not going to develop more these ideas here. But we hope that this rough comparison between the two frameworks has shown that, although there exist some important relevant differences, there are also strong similarities between them.

7 Conclusions

We have reviewed the stit semantic framework and the main stit operators that have been proposed by Belnap, Perloff and Horty (among others). We discussed the problem of how to model the notion of collective agency, both in the sense of a joint action of a group of agents (reviewing some extensions of the stit operators to groups of agents) and as the agency of collective entities, like organisations. We have shown how we can describe 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. Finally, we have shown how we can define 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 formulae of the form [*a*: φ]ψ, with the following informal meaning: ''if agent *a* selects an action that leads to the truth of φ, after

[58]Naturally, φ must also be a transition formula in order to build formula [*a*]φ. But that is what happens when φ is a direct action sentence, which is the case with more interest to us, as we have referred.

such action is performed, ψ is the case'' (or, simply, ''after *a* choosing φ, ψ is be the case''). We ended, comparing our approach with Sergot's recent proposal.

Further technical research is needed on how to implement the *nearly now* requirement, in characterising the logical properties of the modal operators herein proposed, and in integrating these operators with the counts-as, acting in a role and deontic operators, a combination that corresponds to the logical setting that we think it is necessary to characterise organisations and other forms of collective acting and interacting, at an adequate abstract level. A comparison with the approach in [8], and a closer comparison with Sergot's proposal, particularly regarding the various group agency operators proposed in [39], should also be done. Finally, work must also be done in illustrating the use of this logical framework within practical and relevant examples.

Acknowledgements

This paper constitutes an extended and revised version of the paper [12] that was presented at the *3rd International Workshop on Normative Multiagent Systems*, held in Luxembourg, from 15th to 16th of July, 2008. The author wants to thank both the anonymous referees (of [12] and of this revised version) and the participants of that Workshop for their suggestions and criticisms, that have led to the correction and clarification of some aspects of the paper and to improve its quality and effectiveness. The author also wish to thank Mark A. Brown for some helpful comments.

Finally, the author wants to gratefully acknowledge the support from FCT, POCTI-219, FEDER.

References

[1] Belnap, N., Perloff, M. Seeing To It That: A Canonical Form for Agentives. Theoria 54 (1989) 175–199.

[2] Belnap, N. Backwards and forwards in the modal logic of agency. Philosophy and Phenomenological Research 51 (1991) 777–807.

[3] Belnap, N., Perloff, N. In the Realm of Agents. Annals of Mathematics and Artificial Intelligence 9 (1993) 25–48.

[4] Belnap, N., Perloff, M., Xu, M. Facing the Future: Agents and Choices in Our Indeterminist World. Oxford University Press (2001).

[5] Broersen, J., Herzig, A., Troquard, N. From Coalition Logic to STIT. In: Proceedings LCMAS 2005 . Electronic Notes in Theoretical Computer Science, 157 (4), Elsevier (2006).

[6] Broersen, J., Herzig, A., Troquard, N. A STIT-extension of ATL. In: Proceedings Tenth European Conference on Logics in Artificial Intelligence (JELIA'06), Lecture Notes in Artificial Intelligence, Vol. 4160, Springer (2006) 69–81.

[7] Broersen, J., Herzig, A., Troquard, N. Embedding Alternating–time Temporal Logic in Strategic STIT Logic of Agency, Journal of Logic and Computation 16:5 (2006) 559–578.

[8] Broersen, J. A Complete STIT Logic for Knowledge and Action, and Some of Its Applications. In: Proceedings Declarative Agent Languages and Technologies VI, 6th International Workshop DALT 2008, Lecture Notes in Artificial Intelligence, Vol. 5397, Springer (2009) 47–59.

[9] Brown, M.A. Agents with Changing and Conflicting Commitments. In: MacNamara, P., Prakken, H. (eds.), Norms, Logics and Information Systems. Frontiers in Artificial Intelligence and Applications, Vol. 49, IOS Press (1999) 109–125.

[10] Brown, M.A. Acting with an End in Sight. In: Gouble, L., Meyer, J.-J. Ch. (eds.), Deontic Logic and Artificial Normative Systems. Lecture Notes in Artificial Intelligence, Vol. 4048, Springer (2006) 69–83.

[11] Brown, M.A. Acting, Events and Action. In: van der Meyden, R., van der Torre (eds.), Deontic Logic in Computer Science (ΔEON 2008). Lecture Notes in Artificial Intelligence, Vol. 6076, Springer (2008) 19–33.

[12] Carmo, J. Collective Action, Direct Action and Dynamic Operators, In: Boella, G., Pigozzi, G., Singh, M., Verhagen, H. (eds.), Proceedings of the Third International Workshop on Normative Multiagent Systems (NorMAS-08), ISBN 2-919940-48-1 (2008), 31–50.

[13] 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) (2001) 129–163.

[14] Chellas, B.J. The Logical Form of Imperatives. Dissertation, Stanford University (1969).

[15] Chellas, B.J. Modal Logic – An Introduction, Cambridge University Press (1980).

[16] Chellas, B.J. Time and Modality in the Logic of Agency. Studia Logica 51 (1992) 485–517.

[17] Elgesem, D. Action Theory and Modal Logic. Dissertation, University of Oslo (1993).

[18] Elgesem, D. The Modal Logic of Agency. Nordic Journal of Philosophical Logic, Vol. 2, No. 2 (1997) 1–46.

[19] Goranko, V., Jamroga, W. Comparing semantics of logics for multi-agent systems. Synthese 139 (2) (2004) 241–280.

[20] Grossi, D. Designing Invisible Handcuffs (Formal Investigations in Institutions and Organisations for Multi-agent Systems)., SIKS Dissertation Series No. 2007–16 (2007).

[21] Harel, D., Kozen, D., Tiuryn, J. Dynamic Logic. In: Gabbay, D.M., Guenthner, F. (eds.), Handbook of Philosophical Logic, 2nd ed., Vol. 4, Kluwer, Dordrecht (2002) 99–217.

[22] Hilpinen, R. On Action and Agency. In: Ejerhed, E., Lindström, S. (eds.), Logic, Action and Cognition – Essays in Philosophical Logic. Trends in Logic, Studia Logic Library, Vol. 2, Kluwer (1997) 3–27.

[23] Horty, J., Belnap, N. The deliberative stit: A study of action, omission, ability, and obligation. Journal of Philosophical Logic 24, (1995) 583–644.

[24] Horty, J. Agency and Deontic Logic. Oxford University Press (2001).

[25] Jones, A.J.I., Sergot, M. On the Characterisation 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. Wiley (1993) 275–307.

[26] Jones, A.J.I., Sergot, M. A formal characterisation of institutionalized power. Journal of the IGPL 4 (3) (1996) 429–445.

[27] Kanger, S. New Foundations for Ethical Theory, Stockholm (1957) (Reprinted in: Hilpinen, R. (ed.), Deontic Logic: Introductory and Systematic Readings. D. Reidel, (1971) 36–58).

[28] Kanger, S. Law and Logic. Theoria 38 (1972).

[29] Kooi, B., Tamminga, A. Moral Conflicts between Groups of Agents. Journal of Philosophical Logic 37 (2008) 1–21.

[30] Lindahl, L. Position and Change - A Study in Law and Logic. Synthese Library 112, D. Reidel (1977).
[31] 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 (2003) 145–184.
[32] Pörn, I. The Logic of Power. Oxford, Blackwell (1970).
[33] Pörn, I. Action Theory and Social Science: Some Formal Models, Synthese Library 120, D. Reidel (1977).
[34] Santos, F., Carmo, J. Indirect Action, Influence and Responsibility. In: Brown, M., Carmo, J. (eds.), Deontic Logic, Agency and Normative Systems Springer, Workshops in Computing Series (1996) 194–215.
[35] Segerberg, K. Bringing it about. Journal of Philosophical Logic 18 (1989) 327–347.
[36] Sergot, M. Normative Positions. In: P. MacNamara, P., Prakken, H. (eds.), Norms, Logics and Information System: New Studies in Deontic Logic and Computer Science. Amsterdam, IOS Press (1999), 289–308.
[37] Sergot, M. The Language (C/C+)$^{++}$". AlfeBiite Working Document, 2 July 2002 (2002).
[38] Sergot, M. Action and Agency in Norm-Governed Multi-Agent Sistems. In: Artikis, A., O'Hare, G., Stathis, K. (eds.), Proceedings 8th Annual International Workshop "Engineering Societies in the Agents World" (ESAW'07), Lecture Notes in Artificial Intelligence, Vol. 4995, Springer (2008).
[39] Sergot, M. The Logic of Unwitting Collective Agency. Technical Report 2008/6, Department of Computing, Imperial College, London (2008).
[40] Sergot, M., Richards, F. On the Representation of Action and Agency in the Theory of Normative Positions. Fundamenta Informaticae 45 (2001), 1–21.
[41] Singh, M.P. Toward a Model Theory of Actions: How Agents do it in Branching Time. Computational Intelligence, 1998.
[42] Wölfl, S. Review of Nuel Belnap, Michael Perloff, and Ming Xu's *Facing the Future*, Notre Dame Philosophical Reviews (2002). Published at: http://ndpr/icaap.org/content/archives/2002/8/wolfl-belnap.html.
[43] Wölfl, S. Qualitative Action Theory (A Comparison of the Semantics of Alternating-Time Temporal Logic and the Kutschera-Belnap Approach to Agency). In: Alferes, J., Leite, J. (eds.), Proceedings of the Ninth European Conf. on Logics in Artificial Intelligence (JELIA'04), Lecture Notes in Artificial Intelligence, Springer (2004), 70–81.
[44] Xu, M. Causation in Branching Time (I): Transitions, Events and Causes. Synthese 112 (1997) 137–192.

Appendix

- Result SCSR ("S*ame Choices Same Results*"):
  If $\mathsf{M} \models_{h,m} [\text{G } \mathit{dstit}\text{: } \varphi]$ and $h' \in \cap_{a \in G} C_{m,a}(h)$, then $\mathsf{M} \models_{h',m} [\text{G } \mathit{dstit}\text{: } \varphi]$

Proof.
Suppose that (i) $\mathsf{M} \models_{h,m} [\text{G } \mathit{dstit}\text{: } \varphi]$ and that h' is such that (ii) $h' \in \cap_{a \in G} C_{m,a}(h)$ We want to prove (iii) $\mathsf{M} \models_{h',m} [\text{G } \mathit{dstit}\text{: } \varphi]$.

In order to prove the positive condition of (iii), consider any h'' such that $h'' \in \cap_{a\in G} C_{m,a}(h')$

By (ii), we have that (iv) for every $a \in$G, $C_{m,a}(h) = C_{m,a}(h')$.

Thus, $h'' \in \cap_{a\in G} C_{m,a}(h)$, and, by the positive condition of (i), $\mathsf{M} \models_{h'',m} \varphi$, as we wish.

In order to prove the negative condition of (iii), consider any d in G. By the negative condition of (i), it exists some history h'' such that $h'' \in \mathrm{H}_m$ and $h'' \in \cap_{b\in G-\{d\}} C_{m,b}(h)$ and $\mathsf{M} \not\models_{h'',m} \varphi$.

But, by (iv), this implies $h'' \in \mathrm{H}_m$ and $h'' \in \cap_{b\in G-\{d\}} C_{m,b}(h')$ and $\mathsf{M} \not\models_{h'',m} \varphi$, as we wish. ∎

- $\models$[G *dstit*: φ] → [G *dstit*: [G *dstit*: φ]]

PROOF.

Suppose that (i) $\mathsf{M} \models_{h,m}$[G *dstit:* φ]. We want to prove (ii) $\mathsf{M} \models_{h,m}$[G *dstit*: [G *dstit*: φ]].

The negative condition of (ii), follows from the negative condition of (i), since $\mathsf{M} \not\models_{h',m} \varphi$ implies $\mathsf{M} \not\models_{h',m}$[G *dstit*: φ] for any h' (by the validity of schema T).

In order to prove the positive condition of (ii), consider any h' such that $h' \in \cap_{a\in G} C_{m,a}(h)$. We want to prove that $\mathsf{M} \models_{h',m}$ [G *dstit*: φ]. But this follows from (i) taking into account the result SCSR above. ∎

- $\models\neg$[G *dstit*: [G′ *dstit*: φ]], whenever G⊂G′

PROOF (*not* assuming the *something happens* condition):

Suppose, by absurd, that exist M, h and m such that (i) $\mathsf{M} \models_{h,m}$[G *dstit*: [G′ *dstit*: φ]].

By (i) (and the validity of schema T), we have that (ii) $\mathsf{M} \models_{h,m}$ [G′ *dstit*: φ].

Consider $c\in$G′-G. By the negative condition of (ii), it exists some history h' such that

$$h' \in \cap_{a\in G'-\{c\}} C_{m,a}(h) \text{ and } \mathsf{M} \not\models_{h',m} \varphi.$$

But then (since G⊂G′ and $c\in$G′-G),$h' \in \cap_{a\in G} C_{m,a}(h)$, and, by the positive condition of (i):

$$\mathsf{M} \models_{h',m} [\mathrm{G}' \; dstit : \varphi]$$

which implies $\mathsf{M} \models_{h',m} \varphi$, and we get a (the desired) contradiction. ∎

- $\models\neg$[G *dstit*: [G′ *dstit*: φ]], whenever G′⊂G

PROOF (*not* assuming the *something happens* condition):

Suppose, by absurd, that exist M, h and m such that (i) $\mathsf{M} \models_{h,m}$[G *dstit*: [G′ *dstit*: φ]].

By (i) (and the validity of schema T), we have that (ii) $\mathsf{M} \models_{h,m}$ [G′ *dstit*: φ].

Consider $c \in$G-G′. By the negative condition of (i), it exists some history h' such that $h' \in \cap_{a\in G-\{c\}} C_{m,a}(h)$ and $\mathsf{M} \not\models_{h',m}$ [G′ *dstit*: φ]

But then (since G′⊂G and $c \in$G-G′), $h' \in \cap_{a\in G'} C_{m,a}(h)$.

And, from the result SCSR, by (ii), it follows that $\mathsf{M} \models_{h',m}$ [G′ *dstit*: φ], and we get a contradiction. ∎

- $\models\neg$[G *dstit*: [G′ *dstit*: φ]], whenever G and G′ are different

PROOF (assuming the *something happens* condition):

Suppose, by absurd, that exist M, h and m such that (i) $\mathsf{M} \models_{h,m}$[G *dstit*: [G′ *dstit*: φ]].

By (i) (and the validity of schema T), we have that (ii) $\mathsf{M} \models_{h,m}$ [G′ *dstit*: φ].

We have already proved this result for G⊂G′.

So, consider the case where it exists $c\in$G-G$'$ (but without assuming G$'\subset$G).
Then, by the negative condition of (i), it exists some history h' such that

(iii) $h' \in \bigcap_{a\in G-\{c\}} C_{m,a}(h)$ and (iv) $\mathsf{M} \nvDash_{h',m}$[G$'$ *dstit*: φ].

Note that (iii) implies that (v) for every $a\in$G-$\{c\}$, $C_{m,a}(h) = C_{m,a}(h')$

Noting that $c\notin$G$'$ and taking into account (v) (which is relevant in what follows for the possible members common to G and G$'$), by the *something happens* condition, it exists some h'' such that

(vi) $h'' \in (\bigcap_{a\in G} C_{m,a}(h)) \cap (\bigcap_{a\in \mathrm{G}'} C_{m,a}(h'))$

Now, from (vi) it follows that $h'' \in \bigcap_{a\in \mathrm{G}'} C_{m,a}(h')$.

Thus, by result SCSR and (iv), we have (vii) $\mathsf{M} \nvDash_{h'',m}$ [G$'$ *dstit*: φ].

On the other hand, from (vi) it follows that $h'' \in \bigcap_{a\in G} C_{m,a}(h)$.

Thus, by the positive condition of (i), $\mathsf{M} \vDash_{h'',m}$ [G$'$ *dstit*: φ], and a contradiction with (vii) obtains. ■

- Belnap's *strict joint stit* (for the deliberative case) is equivalent to our *joint dstit* operator. That is:
  Defining $\mathsf{M} \vDash_{h,m}$[G *stit:* φ] iff
  (1) $\mathsf{M} \vDash_{h',m} \varphi$ for every history h' such that $C_{m,a}(h, h')$ for every a in G, and
  (2) $\mathsf{M} \nvDash_{h',m} \varphi$ for some history h' that pass through m
  and [G *sstit*: φ] as [G *stit*: φ] $\wedge\forall_{\mathrm{G}'}(\varnothing\neq$G$'\subset$G $\rightarrow\neg$[G$'$ *stit*: φ])
  then $\vDash$[G *dstit*: φ] $\leftrightarrow$ [G *sstit*: φ]

Proof.

a) Proof of $\vDash$[G *dstit*: φ] $\rightarrow$ [G *sstit*: φ]:
Suppose (i) $\mathsf{M} \vDash_{h,m}$[G *dstit*: φ]. We want to prove that $\mathsf{M} \vDash_{h,m}$[G *sstit*: φ].
It is obvious that (i) implies that $\mathsf{M} \vDash_{h,m}$[G *stit*: φ].
Let G$'$ be such that $\varnothing\neq$G$'\subset$G. We want to prove that $\mathsf{M} \nvDash_{h,m}$[G$'$ *stit*: φ].
Let $a\in$G-G$'$. By the negative condition of i), it exists $h' \in \bigcap_{b\in G-\{a\}} C_{m,b}(h)$ such that $\mathsf{M} \nvDash_{h',m} \varphi$.
But $\bigcap_{b\in G-\{a\}} C_{m,b}(h) \subseteq \bigcap_{b\in \mathrm{G}'} C_{m,b}(h)$. Thus $h' \in \bigcap_{b\in \mathrm{G}'} C_{m,b}(h)$ and $\mathsf{M} \nvDash_{h',m} \varphi$.
Thus $\mathsf{M} \nvDash_{h,m}$[G$'$ *stit*: φ] (since the positive condition of $\mathsf{M} \vDash_{h,m}$[G$'$ *stit*: φ] is not satisfied).

b) Proof of $\vDash$[G *sstit*: φ] $\rightarrow$ [G *dstit*: φ]:
Suppose (i) $\mathsf{M} \vDash_{h,m}$ [G *sstit*: φ]. We want to prove that (ii) $\mathsf{M} \vDash_{h,m}$[G *dstit*: φ].
By (i), $\mathsf{M} \vDash_{h,m}$[G *stit*: φ] and so the positive condition of (ii) is satisfied.
If #G=1, the negative condition of (ii) follows from the negative condition of $\mathsf{M} \vDash_{h,m}$[G *stit*: φ].
Suppose #G≠1. In order to prove the negative condition of (ii), let $a\in$G and G$'$=G-$\{a\}$.
By (i), we have that (iii) $\mathsf{M} \nvDash_{h,m}$[G$'$ *stit*: φ].
On the other hand, since (by (i)) $\mathsf{M} \vDash_{h,m}$[G *stit*: φ], it exists $h' \in \mathrm{H}_m$ such that $\mathsf{M} \nvDash_{h',m} \varphi$.
But then, if (iii) is the case, it is because the positive condition of $\mathsf{M} \vDash_{h,m}$[G$'$ *stit*: φ] fails, and so it exists h' such that $h' \in \bigcap_{b\in \mathrm{G}'} C_{m,b}(h)$ and $\mathsf{M} \nvDash_{h',m} \varphi$. But this is what we want to prove. ■

- $\models \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\ \varphi_2]) \leftrightarrow \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2)$

PROOF.

a) Proof of $\models \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\ \varphi_2]) \rightarrow \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2)$:
We have that $\models [a\ cstit\text{:}\ \varphi_2] \rightarrow \varphi_2$ (by the validity of T for *cstit*).
Thus $\models ([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\ \varphi_2]) \rightarrow ([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2)$
And so $\models \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\ \varphi_2]) \rightarrow \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2)$ (as desired).

b) Proof of $\models \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2) \rightarrow \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\ \varphi_2])$:
Suppose, by absurd, that for some M, h and m:
(i) $\mathsf{M} \models_{h,m} \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2)$ and (ii) $\mathsf{M} \not\models_{h,m} \Box([a\ cstit\text{:}\ \varphi_1] \rightarrow [a\ cstit\text{:}\varphi_2])$
Then, by (ii), there exists $h' \in \mathrm{H}_m$ such that
(iii) $\mathsf{M} \models_{h',m} [a\ cstit\text{:}\ \varphi_1]$ and (iv) $\mathsf{M} \not\models_{h',m} [a\ cstit\text{:}\ \varphi_2]$
And, from (iv) it follows that there exists $h'' \in C_{m,a}(h')$ such that (v) $\mathsf{M} \not\models_{h'',m} \varphi_2$.
But (since $h'' \in C_{m,a}(h')$ implies $h'' \in \mathrm{H}_m$) from (i) it follows
(vi) $\mathsf{M} \models_{h'',m} [a\ cstit\text{:}\ \varphi_1] \rightarrow \varphi_2$
And (since $h'' \in C_{m,a}(h')$ implies $C_{m,a}(h'') = C_{m,a}(h')$) from (iii) it follows
$\mathsf{M} \models_{h'',m} [a\ cstit\text{:}\ \varphi_1]$
Thus, by (vi), $\mathsf{M} \models_{h'',m} \varphi_2$, which contradicts (v). ∎