Relationships between actions performed by institutional agents, human agents or software agents

Robert Demolombe[1]

Institut de Recherche en Informatique de Toulouse
France
robert.demolombe@orange.fr

Abstract. The paper presents a logical framework for the integration of interactions between institutional agents, human agents and software agents. It is shown, through a case study, that the relationships between actions performed by these three kinds of agents are defined in terms of the Searle's "counts as" concept and their justifications are based on the roles hold by human agents or by a causal relationships between human agent actions and software agent actions.
The logical framework concentrates on the concepts of counts as, causality and role.

1 Introduction

The interactions between companies and people have been analyzed from a legal point for a long time. More recently their interactions have involved computers, networks and complex pieces of software. For instance, new applications have been developed in the field of electronic commerce where individuals can buy or sale goods to other individuals or to companies. Electronic tools can also be used to organize auctions or to negotiate contracts (see [12, 9, 24]).

This new situation raises basic questions about the legal status of these interactions.

Surprisingly there are quite few norms to regulate them and a possible explanation for this situation may be that some basic concepts, like the legal status of the pieces of software that are used for these interactions, have no sufficiently clear definitions to allow lawyers to define appropriate new norms.

For example, there are research proposals where these pieces of software are defined as "normative agents" (see, for example, [30]) and it is not clear whether this terminology is used just as a metaphor for human beings or, strictly speaking, as legal entities. To make more explicit the risk of confusion we can consider the terminology used in the field of highways regulation, when a norm says, for instance, that "car speed is limited to 130 km/h". The risk of confusion comes from the fact that this norm can be understood as a norm which applies to cars or to their drivers. Nevertheless, for most people it is clear that this norm applies to drivers. However, when we speak about "normative agents" it is not so easy to understand who are the human beings who are the counterparts of the drivers.

The purpose of this paper is to try to clarify these kinds of issues when we consider, from a normative point of view, interacting agents which can be either companies, human beings or pieces of software.

Our approach is to start from the detailed analysis of a case study (section 2) in order to exhibit the key concepts. Then, we propose a logical framework to define clear relationships between the actions performed by the three kinds of agents (section 3) and this framework is applied to the case study to evaluate its appropriateness (section 4). Comparisons with other works are provided in section 5 and in the least section are presented our conclusions and directions for possible further works.

2 A case study

Before to present the case study we define some terms that will be used along this paper. We call "institution" a set of norms in the sense proposed by Searle in [26] For example, the set of norms about international commerce is an institution. Following the same terminology we call "institutional agent" an agent whose existence and legal status is defined in the context of an institution. For example, a company, an hospital, a university or a foundation, where the status defines, for instance, the kind of taxes they have to pay and the legal responsibility of some of their managers.

Let's consider now a case study where a company (the institutional agent I_1) wants to send to another company (the institutional agent I_2) a document doc_1 which is an official price offer to sale a given good, say a computer (this action is called α_1).

Since an institutional agent is an abstract entity it cannot directly do concrete actions like sending a document. However, there are people (for instance the human agent H_1) who hold some given role in the organization of the company I_1 and who can do some concrete actions, like sending a document, which, according to an institution, count as actions performed by the company I_1.

Then, the agent H_1 sends the document doc_1 to I_2. Since the company I_2, like I_1, is an institutional agent, to send send doc_1 to I_2 there must be a norm which defines what are the possible concrete means to send a document to I_2. Let's assume that this means is a mail box which can be accessed by a human agent H_2 who holds some given role in the company I_2 (see figure 1).

According to these norms the fact that H_1 has delivered doc_1 in the mail box (this action is called β_1) and H_2 has accessed the mail box (this action is called β_2) counts as the fact that the company I_1 has sent doc_1 to the company I_2. It is worth noting that H_2's action is not caused by H_1's action even if the former can be influenced by the latter. For example, delivering the document may be registered and a receipt may be required, but even if H_2 is "obliged" to read doc_1 this obligation is not necessarily fulfilled.

In most cases the human agent H_1, instead of bringing himself doc_1 in the mail box, orders another company to bring doc_1 on behalf of himself. Another option, which is what we are interested in in this paper, is to use electronic mail to send a message m_1 to H_2, provided in the institution sending m_1 counts as sending the document doc_1. That is the option we are analyzing in what follows.

In this option H_1 transmits to a computer some pieces of information, in particular the content of the message m_1, the electronic identification of H_2's electronic mail box and possibly other data. After giving this information H_1 clicks on some given field (this action is called γ_1). That causes running a piece of software which is called the software agent S_1 (see figure 1).

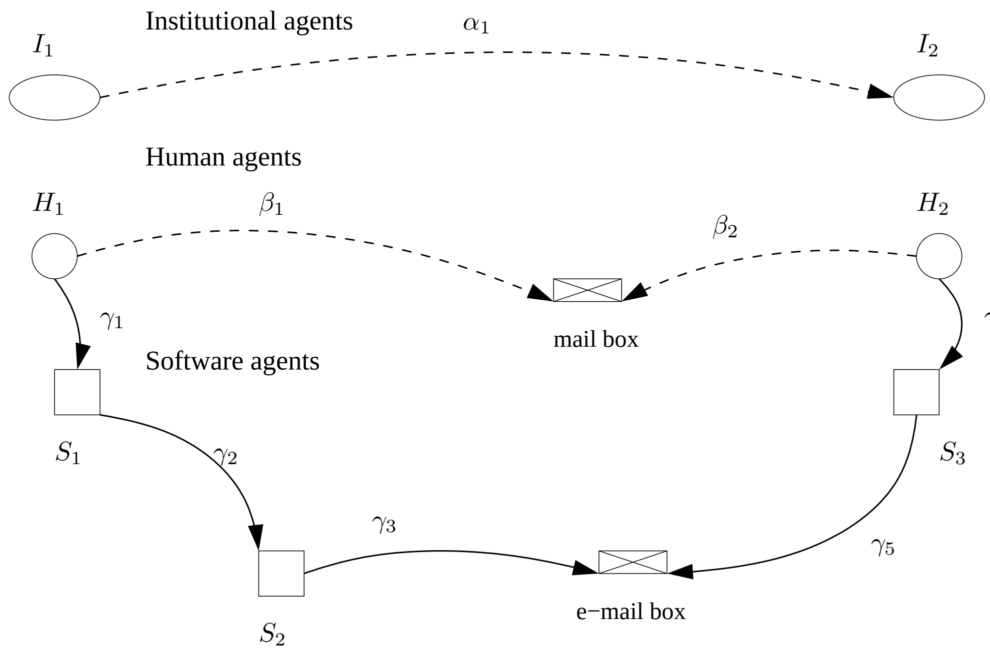

Fig. 1. Relationships between actions.

The software agent S_1 by doing some action γ_2 causes running another software agent S_2 in order to bring the message m_1 in H_2's electronic mail box (this action is called γ_3).

Let's adopt the notation: $agt : act$ to denote the fact that the action act has been performed by the agent agt, and the notation: $(agt_1 : act_1); (agt_2 : act_2)$ to denote the action $agt_1 : act_1$ followed by the action $agt_2 : act_2$.

According to the institution in force in the application domain the sequence of actions $(H_1 : \gamma_1); (S_1 : \gamma_2); (S_2 : \gamma_3)$ counts as the action $H_1 : \beta_1$.

In a similar way the human agent H_2 can access the content of his electronic mail box by clicking on a given field of his computer (this action is called γ_4). That causes running of a software agent S_3 which accesses the mail box and displays the content of the message m_1 on the screen of H_2's computer (this action is called γ_5). As it was pointed out above here also H_2's action is not caused by H_1's action.

According to the institution the sequence $(H_2 : \gamma_4); (S_3 : \gamma_5)$ counts as the action $H_2 : \beta_2$.

Finally, in this context the sequence of actions $(H_1 : \gamma_1); (S_1 : \gamma_2); (S_2 : \gamma_3); (H_2 : \gamma_4); S_3 : \gamma_5)$ counts as the sequence $(H_1 : \beta_1); (H_2 : \beta_2)$, and the sequence $(H_1 : \beta_1); (H_2 : \beta_2)$ counts as the action $I_1 : \alpha_1$.

From this case study we can see that in the context of an institution some actions performed by human agents counts as other actions performed by institutional agents, and some actions performed by software agents counts as actions performed by human agents. The justifications of the relationships between these actions is quite different in the case of an action performed by a human agent and in the case of an action performed by a software agent.

Indeed, in the case of a human agent action the justification is based on the fact that the human agent holds a specific role, while, in the case of a software agent action, it is based on the fact that performance of this action has been caused directly (case of $S_2 : \gamma_3$) or indirectly (case of $S_1 : \gamma_2$) by a human agent action. Then, in the latter case it is a causal relationship which justifies the "counts as" relationship.

Roughly speaking we can say that a software agent is just a means (which can be quite sophisticated) which is used by a human agent in the same way as an elevator can be used to go to some level in a building. There is no fundamental distinction between clicking on some field of a screen and pushing a button of an elevator.

Moreover we can say that a necessary condition (which is not a sufficient condition) for a software agent action to count as a human agent action is that this software agent action has been caused by this human agent action. That is why in the case where the effect of the software agent action is to violate some norm (for instance to inform someone about some private data) the legal responsibility of this violation is ascribed to this human agent.

The previous informal analysis deserves some additional comments.

The first one is that to count as a contribution to the action $I_1 : \alpha_1$ it is not enough that action $H_1 : \beta_1$ was performed by an agent who holds some specific role r_1, it is also necessary that H_1 was acting as an holder of the role r_1. In the following the fact an agent agt does an action act as an holder of the role r will be denoted by: $agt : r : act$.

The second comment is about what has been called a "software agent". In our view a software agent is composed of the concrete representation of a set of instructions and data, and the actions performed by the software agent are the performance of these instructions by a computer. The meaning of these concrete representations are defined by sets of conventions. Finally, it is worth noting that we do not presuppose that software agents have any specific capacities, like "reasoning" capacities, "beliefs", "desires" or "intentions", as most people do in the so called BDI approach of agents (see [22]).

Finally, from the analysis of this case study we can draw the conclusions that the relationships between institutional agent actions, human agent actions and software agent actions are of the kind of "counts as" and that the counts as relationships are justified by the role a human agent holds in the organization of an institutional agent and by the causal relationship between a human agent action and a software agent action. In the next section we shall see how the notions of counts as, role and causality can be formalized in a logical framework.

3 Towards a logical formalization

3.1 Counts as

For the formalization of the counts as operator we have adopted the logical framework proposed in the seminal work of Jones and Sergot [17] and we have only changed minor technical details. In particular, the formalization which is presented here is only defined in the semantics.

In the following it is assumed that ϕ, ψ, θ... denote formulas of a language of propositional logic LP.

We have adopted the following notations.

$\phi \Rightarrow_s \psi$ can be read: ϕ counts as ψ in the institution s.

$D_s\phi$ can be read: ϕ is necessarily true in the institution s or ϕ is recognized by the institution s.

To define the semantics of these two operators a model M is defined as a tuple $M =< W, f_s, d_s, P >$, where W is a set of possible worlds, P is a function which assigns a set of worlds to each atomic formula and which is extended as usual to logical connectives, f_s is a function which assigns in each world a set of sets of worlds to a set of worlds and d_s is a function which assigns to each world a set of worlds. More formally we have:

$f_s \ : \ W \times 2^W \rightarrow 2^{2^W}$

$d_s \ : \ W \rightarrow 2^W$

The satisfiability conditions[1] for these operators are:

$M, w \models \phi \Rightarrow_s \psi$ iff $||\psi||_M \in f_s(w, ||\phi||_M)$

$M, w \models D_s\phi$ iff $d_s(w) \subseteq ||\phi||_M$

where $||\phi||_M$ denotes the set of worlds $\{w \ : \ M, w \models \phi\}$.

Notice that it follows from the second satisfiability condition that D_s is a normal modal operator.

First, we have an inference rule to express the substitutivity of logically equivalent formulas in the antecedent and in the consequent of the counts as operator.

(EQV) If $\models \phi \leftrightarrow \phi'$ and $\models \psi \leftrightarrow \psi'$, then $\models (\phi \Rightarrow_s \psi) \rightarrow (\phi' \Rightarrow_s \psi')$.

[1] See [4] for the satisfiability conditions of normal modal operators and classical modal operators.

Then we have two schemas to express the closure of the consequent and of the antecedent of the counts as operator.

(CC) $\models ((\phi \Rightarrow_s \psi) \wedge (\phi \Rightarrow_s \psi')) \rightarrow (\phi \Rightarrow_s (\psi \wedge \psi'))$

(CA) $\models ((\phi \Rightarrow_s \psi) \wedge (\phi' \Rightarrow_s \psi)) \rightarrow ((\phi \vee \phi') \Rightarrow_s \psi)$

We have also accepted the following transitivity schema.

(S) $\models ((\phi \Rightarrow_s \psi) \wedge (\psi \Rightarrow_s \theta)) \rightarrow (\phi \Rightarrow_s \theta)$

The links between $\Rightarrow_s$ and D_s are expressed by the following schemas:

(D) $\models (\phi \Rightarrow_s \psi) \rightarrow D_s(\phi \rightarrow \psi)$

(C) $\models (\phi \Rightarrow_s \psi) \rightarrow (\phi \rightarrow D_s\phi)$

The intuition of (D) is that the counts as operator entails in the institution s a constraint represented by the material implication, and the intuition of (C) is that if a fact represented by ϕ counts as another fact in s, then this fact is recognized by the institution s as an institutional fact. Then, (C) allows to derive an institutional fact, represented by $D_s\phi$, from a brute fact, represented by ϕ.

The constraints on the functions f_s and d_s which validate the above principles can be found in [17].

3.2 Causality

We want to formalize the fact that the performance of an action which has caused the performance of another action counts as the performance of a third action. In this perspective we refer to actions which have been done. However, to formalize causality we have to consider the states of affairs before the performance of an action and after its performance. Then, we need formal definitions of operators that express that an agent is going to do some action and also of operators that express that an agent has done some action. For that purpose we have defined the following operators whose intuitive meaning is presented below.

$Does_{i:r:\alpha}\phi$: agent i is going to do the action α, acting as holder of the role r, and after performance of α the proposition ϕ holds.

$Done_{i:r:\alpha}\phi$: agent i has performed the action α, acting as holder of the role r, and the proposition ϕ holds.

Since software agent cannot hold any role (from an institutional point of view) we have to define similar operators without reference to a role.

$Does_{i:\alpha}\phi$: agent i is going to do the action α and after performance of α the proposition ϕ holds.

$Done_{i:\alpha}\phi$: agent i has performed the action α and the proposition ϕ holds.

We shall use the notation $Done_{i:r:\alpha}$ (respectively $Done_{i:\alpha}$) to express the fact that action α has been done, without reference to the effects of the action. Then, this notation can be seen as an abbreviation for $Done_{i:r:\alpha}True$ (respectively $Done_{i:\alpha}True$).

These operators allow us to express that some propositions are true after performance of an action. To express that an action has caused that some proposition is true after its performance we need additional operators. These operators and their intuitive meanings are presented below.

$E_{i:r:\alpha}\phi$: agent i has brought it about that ϕ holds by doing the action α and acting as holder of the role r.

$E^{+}_{i:r:\alpha}\phi$: agent i is going to bring it about that ϕ holds by doing the action α and acting as holder of the role r.

$E_{i:\alpha}\phi$: agent i has brought it about that ϕ holds by doing the action α.

$E^{+}_{i:\alpha}\phi$: agent i is going to bring it about that ϕ holds by doing the action α.

To define the semantics of these operators we have defined four families of accessibility relations. They are intuitively presented below.

$wD_{i:r:\alpha}w'$: agent i has started to do the action α in w, acting as holder of the role r, and possibly other actions, and he has ended the action α in w'.

$wD_{i:r:\neg\alpha}w''$: agent i has started to do in w the same actions has he did in w' such that $wD_{i:r:\alpha}w'$, except the action α, and he as ended these actions in w''[2].

The relation $D_{i:r:\neg\alpha}$ is intended to express the counterfactual condition with respect to the relation $D_{i:r:\alpha}$ in the same way as Pörn expresses (see [21] chapter 1, section 5) the counterfactual condition with the relation D' with respect to the relation D[3]. A significant difference is that we have made explicit the worlds where we are before to do the action and after to do the action, while in Pörn's semantics the worlds should be interpreted as histories[4]. Other differences are the fact that agent i is acting as holder of a role and the fact that Pörn does not explicitly mention the name of the action α.

$wD_{i:\alpha}w'$: agent i has started to do the action α in w and possibly other actions, and he has ended the action α in w'.

$wD_{i:\neg\alpha}w''$: agent i has started to do in w the same actions has he did in w' such that $wD_{i:\alpha}w'$, except the action α, and he as ended these actions in w''.

The satisfiability conditions for the operators of the kind *Does* and E^{+} are defined in the same way whatever the agent i is acting as a role holder or not. Then, we use the notation *agt* to denote either: $i : r$ or: i in the following satisfiability conditions.

$M, w \models Does_{agt:\alpha}\phi$ iff $\forall w'(wD_{agt:\alpha}w' \Rightarrow M, w' \models \phi)$

$M, w \models E^{+}_{agt:\alpha}\phi$ iff

(1) $\forall w'(wD_{agt:\alpha}w' \Rightarrow M, w' \models \phi)$ and

(2) $\exists w''(wD_{agt:\neg\alpha}w'' \Rightarrow M, w'' \models \neg\phi)$

The condition (1) expresses that it is sufficient in w that the agent *agt* does α to obtain the effect ϕ and the condition (2) expresses that it was necessary in w that the agent *agt* did α to obtain the effect ϕ.

The semantics of the action operator *Done* (respectively E) is defined in function of the semantics of the action operator *Does* (respectively E^{+})[5].

---

[2] If in w' other agents have performed some actions, it is assumed that in w'' they have performed the same actions as in w'.

[3] The meaning of the counterfactual condition is defined by Pörn as: "*but for i's action it would not be the case that* ϕ".

[4] In [21] the worlds u' related to the world u where we are are defined in that way: "*we must consider all those hypothetical situations* u' *in which the agent does as much as he does in* u", and that is the reason why the accessibility relation D is assumed to be reflexive.

[5] The operators *Done* and E are added to the language because "counts as" statements refer to actions that have been performed.

The intuition of these definitions is that if in a world w we have, for example, $E_{i:\alpha}\phi$, then there must exist a previous world w_1 where the action α has started and where we have $E^+_{agt:\alpha}\phi$. However, the situation may be more complex because the proposition ϕ may also contain an action operator, for example we may have $\phi = E_{j:\beta}\psi$. Then, the meaning of $E_{i:\alpha}(E_{j:\beta}\psi)$ is that in w i has brought it about that j has brought it about that ψ. That means that there must exist a world w_1 where i is going to bring it about that in a further world w_2 j is going to bring it about that in w the proposition ψ holds. That is, in w_1 we have $E^+_{i:\alpha}(E^+_{j:\beta}\psi)$, which has the effect that in w_2 we have $E^+_{j:\beta}\psi$, which itself has the effect that in w we have ψ.

Since there is no fixed limitation in nesting the action operators we have the following recursive definition of the semantics of E and $Done$.

To avoid to repeat similar definitions we adopt the following notations: A_a denotes an action operator which is either $Done$ or E and a is either: $i : r : \alpha$ or: $i : \alpha$, and A^+ denotes $Does$ (respectively E^+) if A denotes $Done$ (respectively E). Then, we have:

$M, w \models A_a\phi$ iff $\exists w_1(\exists w_2(w_1 D_a w_2$ and $Path(\phi, w_2, w))$ and $M, w_1 \models A^+_a T(\phi))$

The formulas denoted by $Path(\phi, w_2, w)$ and $T(\phi)$ are recursively defined in the same way for the operators: $Done_{i:r:\alpha}$, $Done_{i:\alpha}$, $E_{i:r:\alpha}$ and $E_{i:\alpha}$, and these operators can be nested and mixed without any fixed limitation. We have:

- If ϕ is not of the form: $A_a\phi_n$, then
  $Path(\phi, w_n, w) \stackrel{\text{def}}{=} (w_n = w)$ and $T(\phi) \stackrel{\text{def}}{=} \phi$.
- If ϕ is of the form $A_a\phi_n$, then
  $Path(\phi, w_n, w) \stackrel{\text{def}}{=} \exists w_{n+1}(w_n D_a w_{n+1}$ and $Path(\phi_n, w_{n+1}, w))$
  and $T(\phi) \stackrel{\text{def}}{=} A^+_a T(\phi_n)$.

The intuition of the formulas $\exists w_1(\exists w_2(w_1 D_a w_2$ and $Path(\phi, w_2, w))$ is that the sequence of nested actions which have been performed has started in the world w_1, and if there is no nested operator the world w_2 is the world w where we are after a action performance.

It can be easily proved that we have the following properties for the action operators.

(DD) $\models Does_a(Done_a)$

(ED) $\models E^+_a\phi \rightarrow Does_a\phi$

The intuition of (ED) is that if action a causes ϕ, then after a performance ϕ holds.

(DO) $\models Done_a\phi \rightarrow \phi$

(E) $\models E_a\phi \rightarrow \phi$

The intuition of (DO) and (E) is that after a performance ϕ holds. The difference between these operators is that $Done$ does not express causality.

(EE$\wedge$) $\models E^+_a\phi \wedge E^+_a\psi \rightarrow E^+_a\phi \wedge \psi$

Property (EE$\wedge$) is quite intuitive.

(EE$\vee$) $\models E^+_a(\phi \wedge \psi) \rightarrow E^+_a\phi \vee E^+_a\psi$

The intuition of property (EE$\vee$) is that if a causes $\phi \wedge \psi$, then a causes at least ϕ or ψ. If we would not have this property it might happen that action a causes $\phi \wedge \psi$ and a guarantees neither the truth of ϕ nor the truth of ψ.

(EDE) $\models E_a^+\phi \wedge Does_a\psi \rightarrow E_a^+(\phi \wedge \psi)$

The property (EDE) may seem to be counter intuitive, because we are inclined to think that if a causes $\phi \wedge \psi$, then a causes ϕ and a causes ψ, which is not the case in general as shows the following property.

($\neg$EE) $\not\models E_a^+(\phi \wedge \psi) \rightarrow E_a^+\psi$

See, for example, the case where ψ is a tautology. In that case, if we have $E_a^+\phi$, we also have $E_a^+(\phi \wedge \psi)$ and, of course, we don't have $E_a^+\psi$.

3.3 Role

In [21], Pörn proposes this definition: *"Because of their prevalence in normative systems clusters of norms organized in this way deserve a name on their own. We shall call them role structure because in terms of them it is possible to define the sociological notion of a role."* . In other words, if, in a normative system, we repeatedly need to talk about the set of individuals to whom a given set of norms applies, it is convenient to select a name for this set of norms, and the set of norms which is referred to by this name is called a "role".

However, this definition raises a difficulty if the set of norms contains norms defining some kinds of institutional powers. In particular, if an institutional power says that the fact that the role holder has performed some action acting as the role holder counts as something else we see that we refer to the role in a norm which contributes to the definition of the role itself. That is, we have the role name which occurs both in the *definiens* and in the *definendum* and that means that we have a circular definition.

To avoid this circularity problem we shall say that a role refers to a set of norms, but we do not say that a role is defined by a set of norms. In this approach there is no more problem of circular definition.

A situation where, for any given h who holds a given role r in a given institutional agent i in the context of a given institution s, the fact that h has performed a given action α counts as the fact that i has performed a given action β can be represented in a semi-formal way by the following formula.

$\forall h(Holds(h, r, i, s) \rightarrow (Done_{h:r:\alpha}) \Rightarrow_s Done_{i:\beta}))$

where the intuitive reading of the predicate $Holds(h, r, i, s)$ is:

$Holds(h, r, i, s)$: human agent h holds the role r in the organization of the institutional agent i in the context of the institution s.

We insist on the fact that this formula is not an axiom schema, it is an example of norm of the kind counts as, and the symbols r, i, s, α and β are constant symbols. The variable symbol h is universally quantified because this kind of norm is not defined for a specific given individual.

This formula is "semi-formal" because it contains a universally quantified variable h which occurs both as an argument of the predicate $Holds$ and as an index of the modal operator *Done*. To give a formal semantics to this kind of formula is out of the scope of this paper because it would require to go into too long technical details.

The fact that acting of some role holder requires to hold this role and to fulfill some given conditions during the performance of the action can be represented by the following formula.

$\forall h(Done_{h:r:\alpha} \rightarrow (Holds(h,r,i,s) \wedge act.cond_{\alpha,r,i,s}))$
where $act.cond_{\alpha,r,i,s}$ denotes the specific conditions which have to be fulfilled to recognize that the agent has done the action α acting as holder of the role r. Here again r, i, s and α denote constant symbols.

4 Application of the logical framework

We have adopted the following notations to represent some facts in the case study.

ϕ: the price offered by I_1 to I_2 for the good X is Y.

ϕ': there is in H_2's mail box a document doc_1 sent by H_1 whose meaning is represented by ϕ.

ψ': H_2 has got the document doc_1 which is in his mail box.

ϕ'' : H_2 has in his electronic mail box a message m_1 sent by H_1 whose meaning is represented by ϕ.

ψ'': H_2 has read the message m_1 in his electronic mail box.

We assume that we are in the world w of some model M where the following assumptions hold.

It is assumed that H_1 has performed the action formally represented by (1)[6].

(1) $E_{H_1:r_1:\gamma_1} E_{S_1:\gamma_2} Done_{S_2:\gamma_3} \phi''$

The intuitive meaning of (1) is that H_1 has brought it about that the software agent S_1 has brought it about that the software agent S_2 has done an action whose effect is that we have ϕ'' and H_1 was acting as holder of the role r_1.

It is assumed that in the context of the institution s this action counts as the action β_1. Then, we have:

(2) $E_{H_1:r_1:\gamma_1} E_{S_1:\gamma_2} Done_{S_2:\gamma_3} \phi'' \Rightarrow_s E_{H_1:r_1:\beta_1} \phi'$

The intuitive meaning of $E_{H_1:r_1:\beta_1} \phi'$ is that H_1 has brought it about that ϕ' by doing β_1 and acting has holder of the role r_1.

It is also assumed that H_2 has performed the action represented by (3).

(3) $E_{H_2:r_2:\gamma_4} Done_{S_3:\gamma_5} \psi''$

It is assumed that in the context of the institution s this action counts as the action β_2. Then, we have:

(4) $E_{H_2:r_2:\gamma_4} Done_{S_3:\gamma_5} \psi'' \Rightarrow_s E_{H_2:r_2:\beta_2} \psi'$

Finally, it is assumed that the fact that H_1 has performed β_1 and H_2 has performed β_2 counts in the institution s as the fact that I_1 has performed α. Then, we have:

[6] The parenthesis in: $E_{H_1:r_1:\gamma_1}(E_{S_1:\gamma_2}(Done_{S_2:\gamma_3}(\phi'')))$ and in similar formulas are omitted for simplification.

(5) $(E_{H_1:r_1:\beta_1}\phi') \wedge (E_{H_2:r_2:\beta_2}\psi') \Rightarrow_s E_{I_1}\phi$[7].
From (2) and (D) we infer that in w we have:
(6) $D_s(E_{H_1:r_1:\gamma_1}E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi'' \rightarrow E_{H_1:r_1:\beta_1}\phi')$
From (2) and (C) we have:
(7) $E_{H_1:r_1:\gamma_1}E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi'' \rightarrow D_s(E_{H_1:r_1:\gamma_1}E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi'')$
Then, from (1) and (7) we have:
(8) $D_s(E_{H_1:r_1:\gamma_1}E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi'')$
Since D_s is a normal modal operator, from (6) and (8) we have:
(9) $D_s(E_{H_1:r_1:\beta_1}\phi')$
In a similar way from (3) and (4) we can infer that in w we have:
(10) $D_s(E_{H_2:r_2:\beta_2}\psi')$
Since D_s is a normal modal operator, from (9) and (10) we have:
(11) $D_s((E_{H_1:r_1:\beta_1}\phi') \wedge (E_{H_2:r_2:\beta_2}\psi'))$
From (5) and (D) we have:
(12) $D_s((E_{H_1:r_1:\beta_1}\phi') \wedge (E_{H_2:r_2:\beta_2}\psi') \rightarrow E_{I_1}\phi)$
Finally, since D_s is a normal modal operator, from (11) and (12) in w we have:
(13) $D_s(E_{I_1}\phi)$
The intuitive meaning of (13) is that it is recognized by the institution s that I_1 has brought it about that ϕ. It is interesting to point out that to infer (13) we didn't use the transitivity schema (S).

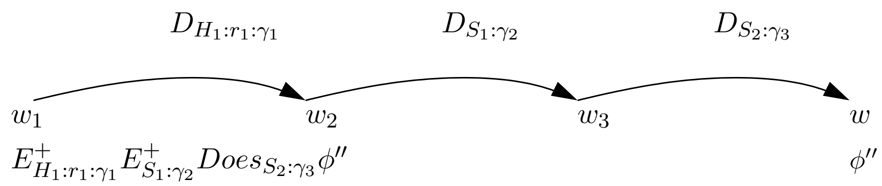

Fig. 2. Interpretation of action operators.

We take opportunity of this case study to show how the satisfiability conditions for the action operator E are evaluated. That is presented for the action: $E_{H_1:r_1:\gamma_1}E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi''$. The general form of this formula is: $A_a\phi_1$, where $A = E$ and $a = H_1 : r_1 : \gamma_1$, and where we have:
$\phi_1 = E_{S_1:\gamma_2}Done_{S_2:\gamma_3}\phi''$
$\phi_2 = Done_{L_2:\alpha_1^3}\phi''$
$\phi_3 = \phi''$

[7] From the semantics of $E_{H_1:r_1:\beta_1}\phi'$ and $E_{H_2:r_2:\beta_2}\psi'$ it is clear that $E_{H_2:r_2:\beta_2}\psi'$ cannot be performed before $E_{H_1:r_1:\beta_1}\phi'$ since the document cannot be accessed in the mail box before it has been stored in it. Nevertheless, there are many other examples where a temporal operator is needed to express that an action has to be performed before another action.

Then, the general form of the satisfiability conditions is instantiated by:

$M, w \models E_{H_1:r_1:\gamma_1}\phi_1$ iff $\exists w_1(\exists w_2(w_1 D_{H_1:r_1:\gamma_1} w_2$ and $Path(\phi_1, w_2, w))$ and $M, w_1 \models E^+_{H_1:r_1:\gamma_1} T(\phi_1))$

From T definition we have:

$T(\phi_1) = E^+_{S_1:\gamma_2} T(\phi_2)$
$T(\phi_2) = Does_{S_2:\gamma_3} T(\phi_3)$
$T(\phi_3) = \phi''$

From $Path$ definition we have:

$Path(\phi_1, w_2, w) = \exists w_3(w_2 D_{S_1:\gamma_2} w_3 \; and \; Path(\phi_2, w_3, w))$
$Path(\phi_2, w_3, w) = \exists w_4(w_3 D_{S_2:\gamma_3} w_4$ and $Path(\phi_3, w_4, w))$
$Path(\phi_3, w_4, w)) = (w_4 = w)$

Then, the final rewriting of these conditions is:

$M, w \models E_{H_1:r_1:\gamma_1} E_{S_1:\gamma_2} Done_{S_2:\gamma_3}\phi''$ iff
$\exists w_1(\exists w_2(w_1 D_{H_1:r_1:\gamma_1} w_2$ and $\exists w_3(w_2 D_{S_1:\gamma_2} w_3$ and $w_3 D_{S_2:\gamma_3} w))$ and
$M, w_1 \models E^+_{H_1:r_1:\gamma_1} E^+_{S_1:\gamma_2} Does_{S_2:\gamma_3}\phi'')$

The sequence of worlds: w_1, w_2, w_3 and w can be seen as the past history that has lead to obtain ϕ'' in w (see figure 2). Indeed, from the satisfiability conditions if we have $E^+_{H_1:r_1:\gamma_1} E^+_{S_1:\gamma_2} Does_{S_2:\gamma_3}\phi''$ in w_1, in w we have $E_{H_1:r_1:\gamma_1} E_{S_1:\gamma_2} Done_{S_2:\gamma_3}\phi''$, and from property (E) we have $E_{S_1:\gamma_2} Done_{S_2:\gamma_3}\phi''$ and $Done_{S_2:\gamma_3}\phi''$ in w. Finally, from property (DO) we have ϕ'' in w.

5 Comparison with other works

There is a limited number of works about the logical formalization of the counts as operator [8] even if some authors have proposed formalizations which are not based on formal logic (see [10].

In [11] Gelati et al. have proposed a variant of Jones and Sergot formalization which is based on a defeasible conditional operator, denoted by $\Rightarrow$, and which can be used for any defeasible normative connection. Then, the counts as connection takes the form: $(\phi \Rightarrow D_s\psi) \wedge (D_s\phi \Rightarrow D_s\psi)$.

Another significant variant has been proposed by Grossi in [13, 14] (a similar approach has been proposed by Lorini et al. in [19]). Grossi has defined several definitions of this operator and all of them share the formal property that the antecedent and the consequent of the counts as operators appear as the antecedent and the consequent of a material implication which is in the scope of modal operators. For example, the "proper classificatory" operator $\phi \Rightarrow^{cl+}_s \psi$[9] is defined as: $[s](\phi \rightarrow \psi) \wedge \neg[u](\phi \rightarrow \psi)$, where the intuition of $[u](\phi \rightarrow \psi)$ is that $\phi \rightarrow \psi$ holds in the context of any institution.

The first unexpected consequence is that, since material implication $\phi \rightarrow \psi$ is logically equivalent to $\neg\phi \vee \psi$, and disjunction is commutative, the antecedent and the consequent have the same "status". That is, there is no distinction

[8] See also in [25] the distinction between the concepts of "counting as" and "emergence" proposed by Sartor.

[9] We have slightly changed the notation to make easier the comparison with Jones and Sergot's definition.

between brute facts and institutional facts, and we can find many examples where $\phi \Rightarrow_s^{cl+} \psi$ entails $\neg\psi \Rightarrow_s^{cl+} \neg\phi$.

Another weakness of this formalization is that from a brute fact ϕ we cannot infer any institutional consequence because brute facts are not in the scope of the modal operator $[s]$. For instance, in our case study, from the brut fact (1) and the counts as assumption (2) it is impossible with these definitions to infer the institutional fact represented by (8).

The logical formalization of causality or agency has deserved a large number of works. A survey of these works can be found in [28] (see also [27]) and [1]. The formalization proposed by von Wright [29] has been a reference for many authors. What has been presented in this paper takes inspiration both from von Wright and from further works by Kanger [18] and Pörn [21]. In [7] Demolombe and Jones have analyzed the relationships between the bringing it about action operators and deontic operators. Hilpinen in [15] has proposed a more refined characterization of the counterfactual conditions.

In [16] Horty and Belnap have defined the so called STIT operators. Their basic idea is that at a given moment an agent can chose among a given set of action options and the option he has chosen causes ϕ iff for all the histories that conform the same choice ϕ is obtained and there exists another choice such that for some history which conforms this latter choice ϕ is not obtained. One of the reasons why we have not adopted this approach is that this formalization is not appropriate to represent situations where agents are software agents whose actions are not the consequence of a deliberative choice in its genuine sense.

In the deliberative STIT approach it is inconsistent[10] to say that an agent i has made a choice whose effect is that an other agent j has made some particular choice. At the opposite, with the bringing it about operator we have defined it is not inconsistent to say, for example, that the software agent S_1 by doing γ_2 has brought it about that the software agent S_2 by doing γ_3 has brought it about that ϕ''.

There are few works about the logical formalization of roles. Cuppens in [5] has proposed a logical formalization where roles are seen as "virtual agents" and, for example, what is permitted to do is what this virtual agent does in some ideal world. In [3, 23, 20] Carmo, Pacheco and Santos have defined a role as a set of obligations, permissions and prohibitions as proposed by Pörn in [21]. In [8] Demolombe and Louis have extended this definition to institutional powers that allow us to create obligations and permissions. Santos and Pacheco in [20] have also defined action operators of the kind to bring it about which are indexed by pairs of agents and roles as we did in this paper though their definition of these operators are closer to Pörn's definition. They also require the constraint that if an agent is acting as holder of a role, then he holds this role. However, as mentioned in section 3.3, we also have to impose additional constraints because even if an agent holds a role he is not acting as holder of this role for every

[10] In [2] it has been formally proved that this inconsistency follows from the independence constraint about agents' choices.

actions he does. These constraints are called "enacting" constraints by Dastani, Dignum and Dignum in [6].

6 Conclusion

The analysis of a case study has allowed us to draw the following general conclusions.

The fact that an action performed by a human agent counts as an action performed by an institutional agent depends on the fact that this human agent holds some particular role in the organization of the institutional agent.

The fact that an action performed by a software agent counts as an action performed by a human agent depends on the fact that the action performed by the software agent has been caused directly or indirectly by this human agent.

The logical framework which has been presented gives definitions in the semantics of the counts as operator and of actions operators of the kind to bring it about to give an account of causality. These actions operators can be nested without any limitation and we can explicitly make the distinction between the fact that an action is going to be performed and the fact that it has been performed. Also, they can be used as well for human agents and software agents.

This framework deserves further works[11] to integrate in a uniform formalism the concepts for role, counts as and causality. Another significant work to be done is the definition of a corresponding axiomatics and to prove its validity and completeness.

References

1. L. Aqvist. Old foundations for the logic of agency and action. *Studia Logica*, 72, 2002.
2. P. Balbiani, A. Herzig, and N. Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.
3. J. Carmo and O. Pacheco. Deontic and action logics for organized collective agency, modeled through institutionalized agents and roles. *Fundamenta Informaticae*, 48:129–163, 2001.
4. B. F. Chellas. *Modal Logic: An introduction.* Cambridge University Press, 1988.
5. F. Cuppens. Roles and Deontic Logic. In A. J. I. Jones and M. Sergot, editors, *Second International Workshop on Deontic Logic in Computer Science*, Oslo, Norway, 1994.
6. M. Dastani, V. Dignum, and F. Dignum. Role-assignement in open agent societies. In *Proceedings of the Second International Conference on Autonomous Agents and Multiagent Systems*, 2003.

[11] The fact that the framework combines classical modalities and normal modalities is not problematic. Indeed, normal modalities can be seen as a particular case of classical modalities. More precisely, if a normal modality is interpreted by the accessibility relation R, its semantics can be represented as well by the function f_R which is defined as follows:
for every w in W: if $R(w) = \Omega_R$, then $\Omega \in f_R(w)$ iff $\Omega_R \subseteq \Omega$.

7. R. Demolombe and A.J. Jones. Actions and normative positions. A modal-logical approach. In D. Jacquette, editor, *Companion to Philosophical Logic*. Blackwell, 2002.
8. R. Demolombe and V. Louis. Norms, institutional power and roles: toward a logical framework. In F. Esposito, Z. W. Ras, D. Malerba, and G. Semeraro, editors, *Foundations of Intelligent Systems*. Springer, LNAI 4203, 2006.
9. M. Esteva, J. Rodríguez-Aguilar, J. Lluis Arcos, C. Sierra, P. Noriega, and B. Rosell. Electronic Institutions Development Environment. In *Proceedings of the 7th international joint conference on Autonomous Agents and Multiagent Systems (AAMAS08)*. 2008.
10. N. Fornara, F. Viganò, and M. Colombetti. Agent communication and institutional reality. In R. van Eijk, M. Huget, and F. Dignum, editors, *Developments in Agent Communication*. Springer Verlag LNAI 3396, 2005.
11. J. Gelati, G. Governatori, A. Rotolo, and G. Sartor. Declarative power, representation, and mandate: A formal anaysis. In T. Bench-Capon, A. Daskalopulu, and R. Winkels, editors, *Frontieres in Artificial Intelligence and Applications, Number 89*. IOS Press, 2002.
12. G. Governatori, M. Dumas, A. H.M. ter Hofstede, and P. Oaks. A formal approach to protocols and strategies for (legal) negotiation. In H. Prakken, editor, *Procedings of the 8th International Conference on Artificial Intelligence and Law*. ACM Press, 2001.
13. D. Grossi. *Desigining Invisible Hancuffs. Formal Investigations in Institutions and Organizations for Multi-Agent Systems*. PhD thesis, Utrecht University, 2007.
14. D. Grossi, J-J. Ch. Meyer, and F. Dignum. The Many Faces of Counts-as: A Formal Analysis of Constitutive Rules. *Journal of Applied Logic*, 6, 2008.
15. R. Hilpinen. On Action and Agency. In E. Ejerhed and S. Lindstrom, editors, *Logic, Action and Cognition: Essays in Philosophical Logic*. Kluwer, 1997.
16. J.F. Horty and N. Belnap. The deliberative STIT: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.
17. A. J. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the Interest Group in Pure and Applied Logics*, 4(3), 1996.
18. S. Kanger. New foundations of ethical theory. In R. Hilpinen, editor, *Deontic logic*, pages 36–58. D. Reidel Publishing Company, 1983.
19. E. Lorini, D. Longin, B. Gaudou, and A. Herzig. The Logic of acceptance: grounding institutions on agents attitudes. *Journal of Logic and Computation*, 19(6), 2009.
20. O. Pacheco and F. Santos. Delegation in a role-based organization. In Alessio Lomuscio and Donald Nute, editors, *Deontic Logic in Computer Science*, LNCS 3065. Springer, 2004.
21. I. Pörn. Action Theory and Social Science. Some Formal Models. *Synthese Library*, 120, 1977.
22. A.S. Rao and M.P. Georgeff. Modeling Rational Agents within a BDI Architecture. In *Proc. 2nd Int. Conf. on Knowledge Representation and Reasoning*. Morgan Kaufmann, 1991.
23. F. Santos and O. Pacheco. Specifying and reasoning with institutional agents. In *Proceedings of ICAIL*, 2003.
24. M. Sardis and G. A. Vouros. Electronic institutions infrastructure for e-chartering. In A. Artikis, G. M. P. O'Hare, K. Stathis, and G. A. Vouros, editors, *ESAW*, LNCS 4995. Springer, 2008.
25. G. Sartor. *Legal Reasoning: A Cognitive Approach to the Law.* Springer, Berlin, 2005.

26. J. R. Searle. *Speech Acts: An essay in the philosophy of language*. Cambridge University Press, New-York, 1969.
27. K. Segerberg. Bringing it about. *Journal of Philosophical Logic*, 18:327–347, 1989.
28. K. Segerberg. Outline of a logic of action. In F. Wolter, H. Wansing, W. de Rijke, and M. Zakharyaschev, editors, *Advances in Modal Logic, Volume 3*. World Scientific Publishing Co., 2002.
29. G. H. von Wright. *Norm and Action*. Routledge and Kegan, 1963.
30. F. Lopez y Lopez, M. Luck, and M. d'Inverno. Normative agent reasoning in dynamic societies. In *Proceedings of the Third International Conference on Autonomous Agents and Multi-Agent Systems*. IEEE Computer Society, 2004.