Towards an ontology of agency and action From STIT to OntoSTIT+

Nicolas TROQUARD [a,b,c], Robert TRYPUZ [b,c] Laure VIEU [a,b]
[a] *Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier & CNRS*
[b] *Laboratorio di Ontologia Applicata, ISTC, CNR, Trento*
[c] *Università di Trento*

Abstract. A variety of disciplines and research areas have separately studied the notions of action, agents and agency, but no integrated and well-developed formal ontology for them is currently available. This paper is a first attempt at bridging this gap, focusing especially on the relationship between agency and action.

The departure point is STIT logic, the most expressive among the current logics of agency. Agency is the relationship between an agent and the states of affairs it brings about, without referring to how this is done, i.e., the actions performed. Since ontological investigations are best done in a first-order framework, making explicit at the language level the domain of quantification, we first propose a first-order theory that is proved equivalent to the propositional modal logic STIT. The domain and language of this theory is then extended to cover actions, obtaining the theory we call OntoSTIT+.

Keywords. ontology of action, agency, action, logic of agency, STIT

Introduction

Action and agency are crucial notions for a variety of application domains, e.g., multiagent systems and interaction modelling, planning and robotics, law and social modelling... Accordingly, many different research areas, among which the quite rich discipline of philosophy of action, have proposed theoretical accounts. Unfortunately, these proposals are often unrelated; a correlate is that no well-developed ontology of action and agency is currently available. This paper is a first attempt at bridging this gap, focusing especially on the relationship between agency and action, mostly studied separately.

STIT logic (in short: STIT) is one of the most suitable logical systems dealing with *agency*, both in terms of expressivity and formal properties. The key idea of agency comes from Anselm around the year 1100, who argued that acting is best described by what an agent brings about or, in STIT terms, "sees to it that" is true. Agency is thus the relationship between an agent (or a group of agents) and the states of affairs it can bring about, without referring to how this is done, i.e., the actions performed. Reducing the ontological commitment is of course positive, but if one wants to reason on actions themselves, considering their preconditions, distinguishing between different ways of reaching a given state of affairs, analysing the internal structure of the action (its participants other than the agent, its way of unfolding in time) and its essential relationship with the agent's mental states, avoiding to introduce actions in the picture becomes impossible.

STIT is a propositional modal logic. Integrating agency and actions in the same framework could be done by extending STIT with some other modal operators dealing more explicitly with actions like those of PDL; this path has begun to be explored in [1]. However, with modal operators, the domains of interest and their ontological properties are not made explicit in the language but left hidden in the models. Another direction is to work directly in the more expressive framework of first-order logic, more suitable to easily formulate many properties and explore the variety of possible ontological choices.

The methodology chosen for the work presented here is therefore to first express the ontological assumptions of STIT in a first-order theory, called OntoSTIT; this is the purpose of Section 2, after a formal presentation of STIT in Section 1. Then, we propose to extend this theory by enlarging its language and its domain of interpretation to include actions proper. Section 3 is thus dedicated to discussing OntoSTIT+. Having started from a decidable modal logic, future work will examine if OntoSTIT+ is suitable as intended models of some extension of STIT that maintains good reasoning properties.

1. STIT logic

This section is a short introduction to STIT, a family of modal logics of agency [2,3]. We start with pointing out the important properties of STIT, which justifies why we have chosen it as a basis. Then we present the language and syntactic structure of this logic as well as its semantics. Doing so, we try to follow the terminology that is used by its authors, although we are aware that some terms used in STIT might be misleading; in such cases we provide clarification.

Formal properties of STIT. STIT is not the only logic of agency, even though it enjoys formal properties that make it particularly attractive. One such property is that STIT is more expressive than two well-known logics of agency, ATL and CL [4,5]. *Alternating-time Temporal Logic* (ATL) is a direct extension of CTL [6] for multi-agent systems, introducing agents and coalitions of agents who can opt, at every state (or 'choice point'), for a particular subset of the possible courses of time [7]. Pauly's *Coalition Logic* (CL) [8] has been introduced independently in game theory to reason about what agents are able to achieve. As shown by Goranko in [9], CL corresponds to the fragment of ATL restricted to some operators. The second important property of STIT is its decidability, proven in [3, Part VI]. This fact makes STIT an appropriate tool for reasoning.

STIT language. In this paper, we focus on the STIT variant based on the operator called Chellas's stit ($cstit$) with many agents. The language of STIT (L_{STIT}) is described as follows: $\phi \triangleq p \mid a = b \mid \neg\phi \mid \phi \wedge \phi \mid \mathbf{F}\phi \mid \mathbf{P}\phi \mid \Box\phi \mid [a\, cstit : \phi]$, where p belongs to a set of atomic propositions $\mathcal{A}tm$ ($p \in \mathcal{A}tm$) and a, b are elements of set of agents $\mathcal{A}gt$ ($a, b \in \mathcal{A}gt$). **F** and **P** are the standard Prior-Thomason's future and past temporal operators. $\Box$ is the historical necessity operator. $[a\, cstit :\ \phi]$ is the agentive operator "agent a sees to it that ϕ".

STIT Models. Before describing the standard STIT models we need to introduce a few concepts regarding the underlying temporal structures. A *branching time frame* is a structure $\langle Mom, < \rangle$ in which Mom is a nonempty set of moments, and $<$ is a transitive and irreflexive partial order relation such that there is no backward branching and every two moments have a common lower bound. In such a branching time frame moments are

ordered in a tree-like structure, where forward branching represents the *indeterminacy* of the future and the very possibility of agency, and the lack of backward branching represents the determinacy of the past.

A maximal set of linearly ordered moments from Mom is a *history*. "Intuitively, each history represents some complete temporal evolution of the world, one possible way in which things might work out" [2]. As a matter of fact, many possible courses of the world are possible, which exactly expresses the idea of indeterminacy. A given moment might be contained in several different histories. Let thus $H_m = \{h | m \in h\}$ be the set of histories passing through m, those histories in which m occurs.

A STIT model is a structure $\mathcal{M}$ of the form $\langle Mom, <, \mathcal{A}gt, Choice, v\rangle$, where $\langle Mom, <\rangle$ is a branching-time frame, v is a valuation function $v : \mathcal{A}tm \rightarrow 2^{Mom \times Hist}$, where $Hist$ is the set all histories. $\mathcal{A}gt$ is a non empty set of agents acting in time (all intentional components are ignored). $Choice : \mathcal{A}gt \times Mom \rightarrow 2^{2^{Hist}}$ is a function whose values are noted $Choice_a^m$ for given agent a and moment m. $Choice_a^m$ is a partition into equivalence classes of the set of histories H_m through m.

Intuitively, the function $Choice$ represents the possible constraints that an agent is able to exercise upon the course of events at a given moment, i.e. the choices open to the agent at that moment, implicitly corresponding to his or her possible actions.[1] By choosing one choice cell, the agent can rule out the other histories that do not belong to this choice cell and that are possible at the moment of her or his choice. Formally, by choosing—or 'acting'—at m, the agent a selects a particular set of histories from $Choice_a^m$ within which the history to be realized then lies. Given a history $h \in H_m$, $Choice_a^m(h)$ represents the particular choice (set of actions) from $Choice_a^m$ containing h. Histories belonging to a particular choice cell are the *possible outcomes* that might result from performing some *underlying action*.

Choices must be effective. The choice available to an agent at a given moment should not allow a distinction between histories that do not branch at that moment. For each agent, any two histories that are undivided at m must belong to the same choice cell of the partition $Choice_a^m$.

Finally, if there are multiple agents, agents' choices must be independent and compatible. For each moment and for any possible choice of each agent a at that moment, the intersection of all the possible choices selected must contain at least one history.

Semantics. Assuming a STIT model $\mathcal{M}$, we can define the conditions of satisfaction in $\mathcal{M}$ for STIT's formulae, starting with standard operators. In the following, m/h is an *index,* i.e., a pair consisting of a moment m in Mom of $\mathcal{M}$ and a history h from H_m, and v is the evaluation function of $\mathcal{M}$.

$\mathcal{M}, m/h \models p$	$\iff$	$m/h \in v(p), p \in \mathcal{A}tm$.
$\mathcal{M}, m/h \models \neg\phi$	$\iff$	$\mathcal{M}, m/h \not\models \phi$
$\mathcal{M}, m/h \models \phi \wedge \psi$	$\iff$	$\mathcal{M}, m/h \models \phi$ and $\mathcal{M}, m/h \models \psi$
$\mathcal{M}, m/h \models \mathbf{P}\phi$	$\iff$	there is some $m' \in h$ s.t. $m' < m$ and $\mathcal{M}, m'/h \models \phi$
$\mathcal{M}, m/h \models \mathbf{F}\phi$	$\iff$	there is some $m' \in h$ s.t. $m < m'$ and $\mathcal{M}, m'/h \models \phi$

Historical necessity (or settledness) at a moment m (in a history h) is defined as truth in all histories passing through m. Formally: $\mathcal{M}, m/h \models \Box\phi \iff \mathcal{M}, m/h' \models$

[1] There are no actions in STIT, but an action can be seen as corresponding to the choice that the proposition denoting its effects is true (now or at some future point) in a selected set of histories.

ϕ for all $h' \in H_m$. When $\Box\phi$ holds at m, p is said to be *settled true at* m. $\Diamond p$ is defined in the usual way as $\neg\Box\neg\phi$, and stands for historical possibility. The intuitive idea is that $\Box\phi$ should be true at some moment if ϕ is true at that moment no matter how the future will turn out.

The extension of a formula is given by: $|\phi|^{\mathcal{M}}_m = \{h \in H_m | \mathcal{M}, m/h \models \phi\}$. $|\phi|^{\mathcal{M}}_m$ is the set of histories h passing through moment m such that the sentence ϕ is true at m/h.

Now we are ready to define the agentive operator $[\_\, cstit: \_]$.[2] Let a be an agent in $\mathcal{A}gt$ and m/h an index, $\mathcal{M}, m/h \models [a\, cstit: \phi] \iff Choice^m_a(h) \subseteq |\phi|^{\mathcal{M}}_m$. A statement of the form $[a\, cstit: \phi]$, expressing the idea that the agent a sees to it that ϕ, is defined as true at an index m/h just in the case the action performed by a (the choice of a) at that index guarantees the truth of ϕ. The action might result in a variety of possible outcomes, but the statement ϕ must be true in each of them, even though the agent cannot determine which one it will be. For example, my action of *buttering the toast* leads to the state that the *toast is buttered*. This state of affair has to hold in all histories belonging to my choice cell, if I want to truly say that *I saw to it that the toast is buttered*. However many other states of affairs may hold in the histories where the *toast is buttered*. For example the *toast is buttered* may lie either in the history where *the toast is buttered and my tea is cold* or in the history where the *toast is buttered and my tea is hot* and so forth.

[3, p. 435-450] provides a sound and complete axiomatization with respect to the class of models $\mathcal{M}$ and proves its decidability.

Axiomatics. The STIT version considered here is axiomatized as follows:

(A0) Axioms for propositional logic
(A0') Axioms for Prior-Thomason's temporal operators **P**, **F**
(A1) S5 axioms for both modal operators $\Box$ and $[\_\, cstit: \_]$
(A2) $\Box\phi \rightarrow [a\, cstit: \phi]$
(A3) Axioms for standard identity in $\mathcal{A}gt$
(AIA$_k$) $\mathit{diff}(a_0, ..., a_k) \wedge \Diamond[a_0\, cstit: p_0] \wedge ... \wedge \Diamond[a_k\, cstit: p_k] \rightarrow \Diamond([a_0\, cstit: p_0] \wedge ... \wedge [a_k\, cstit: p_k])$ ($k \geqslant 1$), where:
(DA) (Distinct agents) $\mathit{diff}(a_0) \triangleq \top$, $\mathrm{diff}(a_0, ..., a_{n+1}) \triangleq \mathit{diff}(a_0, ..., a_n) \wedge a_0 \neq a_{n+1} \wedge ... \wedge a_n \neq a_{n+1}$, for any $n \geqslant 0$

and takes as rules of inference *modus ponens* and *necessitation*:

(RN) from ϕ infer $\Box\phi$

2. STIT Ontology of Agency - OntoSTIT

2.1. A modal or an ontological approach?

As explained in the beginning of last section, STIT is a quite expressive logic of agency. It has very important formal properties, and accordingly, it knows a growing influence. However, from the ontological point of view, it is not totally clear to what extent STIT captures the intuitions of agency, and how this relates to the notion of action, in particular as it is studied in the philosophy of action.

[2] STIT models allow the definition of many stit operators. For instance, $dstit$ (deliberative stit) and $astit$ (achievement stit) are other well-known operators in the STIT literature. Here, we focus only on $cstit$, but $dstit$ can be easily defined by means of $cstit$ and historical necessity ($[a\, dstit: \phi] \triangleq [a\, cstit: \phi] \wedge \neg\Box\phi$).

It is well known that propositional modal logic has expressivity limitations in comparison with first order logic; this is actually why it has better calculability properties. But whereas the latter enables the expression of rich theories capturing almost all intuitions, the former forces us to tie our intuitions into an at times uncomfortable suit. In this sense it is not surprising that inside the ontological community those who deal with the concept of action and agency have little or no interest in STIT Logic, as Belnap, one of the authors of STIT Logic, complained:

> *The modal logic of agency is not popular. Perhaps largely due to the influence of Davidson (see the essays in Davidson 1980 [10]), but based also on the very different work of such as Goldman 1970 [11] and Thomson 1977 [12], the dominant logical template takes an agent as a wart on the skin of an action, and takes an action as a kind of event. This 'actions as events' picture is all ontology, not modality, and indeed, in the case of Davidson, is driven by the sort of commitment to first-order logic that counts modalities as Bad.* [3]

On the other hand—the argument goes—STIT is philosophically well motivated and *"has the advantage that it permits us to postpone attempting to fashion an ontological theory, while still advancing our grasp of some important features of action..."* [3].

Although, as said earlier, is it true that the first-order framework is more adequate to ontological studies, we would like to draw a slightly different picture from Belnap's. As any representation framework, propositional modal logics do carry ontological assumptions, even though these are often hidden in properties of their models rather than explicitly stated in the language. So, even though the focus in STIT work has (deliberately) not been put on ontological questions, STIT is already in some sense an ontology of agency.

In order to clarify what are STIT's ontological assumptions and establish a base ground on which to build a richer ontology of agency and action, we will in the next sections extract those features of action captured by STIT, and make them explicit in a first-order theory proved equivalent to it, that we will call *OntoSTIT*.

2.2. From STIT to OntoSTIT

We first present the new first-order language we will be using, and then the axiomatic theory that we call OntoSTIT. Following the technique of 'T-encoded semantics' [13,14], and thanks to STIT's completeness with respect to the class of models $\mathcal{M}$ (see above), it can be shown that STIT is equivalent to OntoSTIT.

2.2.1. Language

OntoSTIT is a theory of first-order logic with identity and its language, $L_{OntoSTIT}$, is defined in a standard way. We nevertheless assume the following conventions for variables and constants symbols of $L_{OntoSTIT}$:

- Ω is the set of variables ranging on Particulars: $x_1, ..., x_n (..., s, t, x, x', x'', y, z, ...)$;
- Λ is the set of constants denoting Particulars: **a**, **h**, **m**, **y**, **z**, ...;
- Π is the set of constants denoting States of Affairs: **p**, **p'**, **p"**, ...;
- Δ is the set of constants denoting (primitive) Universals: AG, MO, HT, IN, PRE, $HOLDS$, PO.

The latter predicate constants are understood as, respectively, "is an agent", "is a moment", "is a history", incidence between a moment and a history, precedence between moments, the relation such that at a moment and a history a proposition "holds", the relation such that an agent at a moment makes sure that two histories are both "possible outcomes" of its action.

The models of OntoSTIT are those of STIT, the class of models $\mathcal{M}$. The domain of quantification in which Ω is interpreted covers agents, moments and histories. Even if this is a first-order theory, we need to refer to propositions. The language contains therefore a set of constants that could be seen as denoting *reified* atomic propositions, but will simply be interpreted as states of affairs in $\mathcal{M}$. The truth of such propositions is asserted exclusively via the (meta-)predicate $HOLDS(m, h, \mathbf{p})$ which expresses the idea of STIT that the proposition **p** is true at the moment m and the history h. No Boolean or modal combination of these propositions is allowed within $HOLDS$.

2.2.2. Characterization of primitive relations and categories; definitions

Order on moments. The precedence relation PRE between moments (As1) is transitive (As2) and irreflexive (As3). The linearity in the past is expressed by (As4). (As5) says that any two moments have a lower bound (historical connection).

(As1) $PRE(x, y) \rightarrow MO(x) \wedge MO(y)$[3]
(As2) $PRE(x, y) \wedge PRE(y, z) \rightarrow PRE(x, z)$
(As3) $\neg PRE(x, x)$
(As4) $PRE(x, z) \wedge PRE(y, z) \rightarrow x = y \vee PRE(x, y) \vee PRE(y, x)$
(As5) $\exists z((PRE(z, x) \vee z = x) \wedge (PRE(z, y) \vee z = y))$

Agents. We assume that there is at least one agent (As6). Nothing more is known about agents in OntoSTIT, just as in STIT.

(As6) $\exists x AG(x)$

Moments and histories. There is at least one moment (As7). In STIT models, a history is a set of moments and the relationship between a moment and a history is expressed by $m \in h$. In OntoSTIT language, a history is denoted by a particular individual and no set theoretical axioms are assumed. We simply express the relation between moments and histories by the relation $IN(x, y)$: "the moment x is *in* the history y" or "the history y passes through the moment x" (As8). For any moment, there is some history that passes through it (As9). (As10) is an axiom schema ensuring that when a proposition **p** holds at the moment x and the history y, x is in y.

(As7) $\exists x MO(x)$
(As8) $IN(x, y) \rightarrow MO(x) \wedge HT(y)$
(As9) $MO(x) \rightarrow \exists y IN(x, y)$
(As10) $HOLDS(x, y, \mathbf{p}) \rightarrow IN(x, y)$, for each constant **p** in Π

[3]Universal quantifications over whole formulas are left implicit. We make use of the standard priorities between connectives to avoid unnecessary bracketing.

Histories. That histories denote maximally linearly ordered sets is guaranteed by axiom (As11), using a defined predicate *MLO* for maximally linearly ordered (Ds2), itself based on the defined predicate *LO(x)* for linearly ordered (Ds1). Theorem (Ts1) expresses the idea that if the same moments are in two histories then those histories are identical. The predicate UD, for undivided, can be defined: two histories x and y are undivided at moment z if and only if for some moment t later than z, it is the case that t is in x and y.

(Ds1) $LO(z) \triangleq \forall x, y(IN(x,z) \wedge IN(y,z) \rightarrow x = y \vee x < y \vee y < x)$
(Ds2) $MLO(x) \triangleq LO(x) \wedge \neg\exists y(LO(y) \wedge x \neq y \wedge \forall z(IN(z,x) \rightarrow IN(z,y))$
(As11) $HT(x) \rightarrow MLO(x)$
(Ts1) $\forall z(IN(z,x) \leftrightarrow IN(z,y)) \leftrightarrow x = y$
(Ds3) $UD(x,y,z) \triangleq \exists t(PRE(z,t) \wedge IN(t,x) \wedge IN(t,y))$

Possible Outcome. The predicate $PO(x,y,z,t)$, for possible outcome, expresses the intuitions that are behind the $Choice$ function in STIT: at moment y, histories z and t – that pass through y (As13)[4] – are the possible outcomes of some action performed by agent x (As12) (see Figure 1). We call the histories z and t 'possible outcomes' because each of them *might* result from the action performed by the agent x at y although he cannot determine which will be the actual one. In other words, an agent by his action restricts the possible futures to those histories that are possible outcomes of his action. Note that as STIT, OntoSTIT does not explicitly model action. In other words, actions are not present as individuals in our ontology. That is why we cannot express the intuition, neither in STIT nor in OntoSTIT, that an agent performs a particular action. However this will be possible in OntoSTIT+ (see Section 3).

(As12) $PO(x,y,z,t) \rightarrow AG(x) \wedge MO(y) \wedge HT(z) \wedge HT(t)$
(As13) $PO(x,y,z,t) \rightarrow IN(y,t)$

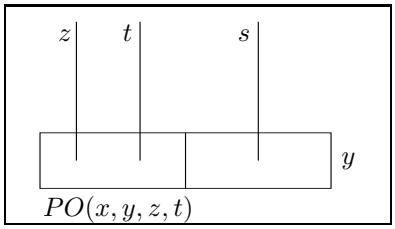

Figure 1. At the moment y, the histories z and t are the possible outcomes of some action performed by the agent x.

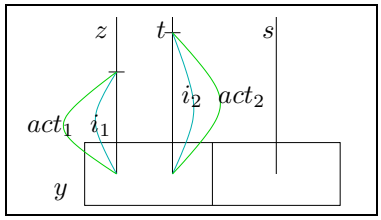

Figure 2. Action Token act_1 (act_2) Lies-On History z (t). Action Token act_1 (act_2) Runs-Through Interval i_1 (i_2).

Considering that the first two arguments are fixed, PO is an equivalence relation. It is reflexive (As14), transitive (As15) and symmetric (As16):

(As14) $IN(y,z) \rightarrow PO(x,y,z,z)$
(As15) $PO(x,y,s,t) \wedge PO(x,y,t,z) \rightarrow PO(x,y,s,z)$
(As16) $PO(x,y,z,t) \rightarrow PO(x,y,t,z)$

Axiom (As17) says that histories that are undivided at moment y are possible outcomes of the same action.

[4]Because of (As16), in axiom (As13) we do not need to explicitly write that also $IN(y,z)$.

(As17) $PO(x, y, t, t') \wedge UD(t', t'', y) \rightarrow PO(x, y, t, t'')$

The axiom schema (As18) expresses the independence of choices. It means that at each moment y there is at least one history t that is common to all agents' possible choices.

(As18) $PO(x_1, y, z_1, t_1) \wedge ... \wedge PO(x_k, y, z_k, t_k) \rightarrow \exists t(PO(x_1, y, z_1, t) \wedge ... \wedge PO(x_k, y, z_k, t))$ for any $k > 1$

2.2.3. Equivalence between STIT and OntoSTIT

To prove that STIT and OntoSTIT are equivalent, we use the technique of 'T-encoded semantics' [13,14], using a function $T_{\dot{x},\dot{y}}$ that enables us to translate formulae of STIT language into formulae of OntoSTIT. This is mainly routine.

Equivalence theorem

For all ϕ in L_{STIT}, ϕ is theorem of STIT iff $\forall x \forall y T_{\dot{x},\dot{y}}(\phi, \{x\}, \{y\})$ is a theorem of OntoSTIT, with x and y being new variables, and the interpretation of $L_{OntoSTIT}$ being constrained s.t. $\omega(x) = \dot{x}$, $\omega(y) = \dot{y}$, where ω transforms variables ranging on Ω into agents, moments or histories of a STIT model $\mathcal{M}$.

2.2.4. How to express agency in OntoSTIT?

The idea of agency is expressed in OntoSTIT by two concepts: possible outcome (PO) and the predicate $HOLDS$ on effects of choice/action. This means that actions themselves are not present in our first-order theory. We can express in OntoSTIT that an agent saw to it that some state of affairs holds (e.g. *the light is off*), even though we still cannot explicitly say by means of which action he/she has done it (we cannot make sure that *the agent switched off the light* rather than *the agent unscrewed the bulb*).

Consider the instantaneous action of *switching off the light* performed by $Robert$, now. We need to be sure that in all possible outcomes of this action it is the case that the *light is off* (we assume that the actual moment is named **n** and the actual history **h**):

(Es1) $\forall h(PO(Robert, \mathbf{n}, \mathbf{h}, h) \rightarrow HOLDS(\mathbf{n}, h, \textit{Light is off}))$

What is more, we want to say that *Robert switches off the light* is true only if *the light was on* just before the action was performed:

(Es2) $\forall x PRE(x, \mathbf{n}) \rightarrow \exists y(PRE(x, y) \wedge PRE(y, \mathbf{n}) \wedge \neg HOLDS(y, \mathbf{h}, \textit{Light is off})) \wedge \forall h(PO(Robert, \mathbf{n}, \mathbf{h}, h) \rightarrow HOLDS(\mathbf{n}, h, \textit{Light is off}))$.

In OntoSTIT (as in STIT) we can also express the idea that an agent brought about some state of affair but he could not have done it or simply it could have happened that that state of affair does not hold. For example we say that *Robert switches off the light*, now, but also that *the light might have been still on*.

(Es3) $\forall h \ (PO(Robert, \mathbf{n}, \mathbf{h}, h) \rightarrow HOLDS(\mathbf{n}, h, \textit{Light is off})) \wedge \exists s(IN(\mathbf{n}, s) \wedge \neg HOLDS(\mathbf{n}, s, \textit{Light is off})$[5]

[5] In STIT this formula can be expressed as follows: $[Robert\ cstit: \textit{Light is off}] \wedge \neg\Box(\textit{Light is off})$ which is equivalent to the formula $[Robert\ dstit: \textit{Light is off}]$. From now on we do not include the preconditions in formulas representing actions.

Notice that in (Es1), (Es2) and (Es3) the moment of choice, **n**, and the moment in which the effect of the action (*the light is off*) comes out, are the same. This expresses the assumption that the action of *switching off the light* is punctual or instantaneous. Instantaneity tightly binds the outcome of the action to the choice of performing that action. Nevertheless it is possible to separate the moment of choice, **n**, and the moment of appearance of the outcome, m.

Let's consider *swimming* and the specific action *Robert swims from point A to point B*. This action belongs to the group of actions that do not go beyond bodily movement.[6] In STIT, Robert's action is expressed by the sentence: *at point A, Robert sees to it that he will be in point B* which, if true, means that at the moment of the choice, when Robert is in point A, Robert is guaranteed to reach point B. This is because all actions (or rather, all choices) are successful in STIT. This seems far too strong an assumption, as in real life, agents do change their minds and actions can abort. There is thus in STIT an agentive gap between the choice and the effects.

A similar problem occurs in the case of actions that do go beyond bodily movement, as for example with *Booth's killing of Lincoln*, (Es4), by shooting him [16]:

(Es4) $\forall h(PO(Booth, \mathbf{n}, \mathbf{h}, h) \rightarrow \exists m(PRE(\mathbf{n}, m) \wedge HOLDS(m, h,$*Lincoln is dead*$))$

(Es4) is the translation of the STIT formula: $[Booth\ cstit\colon$ **F**(*Lincoln is dead*)$]$. Between the moment when *Booth chooses to kill Lincoln* and the moment when *Lincoln is dead*, we have a temporal gap. And we still have the inadequate assumption in STIT that the action consisting of the sequence of events – Booth pulling the trigger, the bullet flying, the bullet entering Lincoln, Lincoln dying – is fully determined by Booth's choice. This means that between the start of the action and the moment when its effect appears, the action cannot be stopped, neither for reasons internal to the agent (which in this case is impossible if we assume the pulling the trigger is instantaneous) nor for any external forces. The temporal gap is here both an agentive and a causal gap.

STIT's assumption that actions are always successful corresponds to the fact that actions are seen *ex post acto*. It is thus in some sense deliberate that only actions that have succeeded are taken into account[7]. As we have seen there are nevertheless good reasons to take a different point of view on actions. Indeed, this is why an extension to STIT has been proposed in [17], to include the new operator "is seeing to it that".

The *ex post acto* view solves the problem of the possible gap between the choice and the action's outcome by simply assuming some kind of determinism of choice, and [17] solves it by assuming the existence of default 'strategies'. OntoSTIT obviously inherits the undesired properties of STIT. To follow more closely findings in philosophy of action, we claim that we should avoid the agentive gap by representing explicitly the persistence of the agent's choice (intention) till the end of the action. Adding the possibility to directly refer to actions is therefore an obvious solution, which moreover opens the path for yet other extensions aimed at accounting for the richness of action concept. The extension of OntoSTIT to actions is the subject of the next section.

[6] Searle [15] claims that no action goes beyond bodily movement. Here we do not take issue on this.

[7] The formula $[a\ cstit\colon \phi] \rightarrow \phi$ is theorem of STIT.

3. Towards an Ontology of Action - OntoSTIT+

In this section we show how OntoSTIT might be extended with actions, obtaining the new theory OntoSTIT+. Its intended models extend the domain of class $\mathcal{M}$ with actions and intervals. We distinguish between action tokens and their 'action courses', which are the different possible ways a single action (i.e., an action token) might unfold in time along different histories. We show at the end of this section that in OntoSTIT+, the problems just described are solved.

3.1. Language.

The language of OntoSTIT+ is that of OntoSTIT extended with new universals. Let Δ_+ be the set of all explicitly introduced universal of OntoSTIT+, $\Delta_+ = \Delta \cup \{INT, ACT, Act, INI, CO, RT, LON, AGO\}$. These new predicate constants are understood as, respectively, "is an interval", "is an action token", "is an action course","a moment is in an interval", "an action course is a course of an action token", "an action course runs through an interval","an action course lies on a history" and "an agent is the agent of an action course".[8]

3.2. Characterization of categories and primitive relations; definitions

Intervals. INI relates moments and intervals (Ap1). All intervals are linearly ordered (Ap2). (Dp1) and (Dp2) define beginning and end of intervals. Any interval has a beginning and an end (Ap3). The unicity of beginning and end for each interval is guaranteed by (Dp1), (Dp2) and (Ap2). Intervals are convex (Ap4). It is worth noting that nothing prevents a beginning of an interval from being equal to its end, so degenerated intervals are possible. (Dp3) defines the relation of temporal part between an interval and a history. For each interval there is a history of which it is temporal part (Ap5). However an interval may belong to more than one history (non-unicity).

(Ap1) $INI(x,y) \rightarrow MO(x) \wedge INT(y)$
(Ap2) $INT(x) \rightarrow \forall x,y(INI(x,z) \wedge INI(y,z) \rightarrow x = y \vee PRE(x,y) \vee PRE(y,x))$
(Dp1) $BEG(x,y) \triangleq INI(x,y) \wedge \forall z(PRE(z,x) \rightarrow \neg INI(z,y))$
(Dp2) $END(x,y) \triangleq INI(x,y) \wedge \forall z(PRE(x,z) \rightarrow \neg INI(z,y))$
(Ap3) $INT(x) \rightarrow \exists y,z(BEG(y,x) \wedge END(z,x))$
(Ap4) $INT(x) \wedge INI(k,x) \wedge INI(l,x) \wedge PRE(k,y) \wedge PRE(y,l) \rightarrow INI(y,x)$
(Dp3) $TP(x,y) \triangleq \forall z(INI(z,x) \rightarrow IN(z,y))$
(Ap5) $INT(x) \rightarrow \exists y(TP(x,y))$

Actions. The relation RT binds an action course to an interval (Ap6). The time of each action course is always fixed: there is exactly one interval such that it runs through it (Ap7). The predicate $CO(x,y)$ links an action course to an action token (Ap8). For each action course there is *exactly one* action token it it is a course of (Ap9). Similarly, for each action token there is *at least one* action course which is a course of it (Ap10). (Ap11) and theorem (Tp1) say that for each action token (and all its courses) we can always find exactly one agent that is agentive for it. (Dp4 - Dp6) define the predicates: $BAct(x,y)$, $EAct(x,y)$ and $BACT(x,y)$ which should be understood respectively as "moment x is

[8] In a larger setting such as DOLCE, AGO would be subsumed by participation.

a beginning of action course y", "moment x is an end of action course y", and "moment x is a beginning of action token y". The unicity of beginning and end of each action course is guaranteed by the unicity of the interval of each action course (Ap7) and the unicity of beginning and end for each interval (Dp1, Dp2, Ap2). (Ap12) guarantees that all action courses of the same action token have the same starting moment, even though they may have different ends. This is why the unicity of an action token's end (that we do not define) cannot be guaranteed, whereas its beginning exists and is unique. Finally, we define the predicate $LON(x, y)$ for "the action course x lies on the history y" by: there is an interval s such that x runs through it and s is a temporal part of y (Dp7).

(Ap6) $RT(x, y) \rightarrow Act(x) \wedge INT(y)$
(Ap7) $Act(x) \rightarrow \exists! y(RT(x, y))$
(Ap8) $CO(x, y) \rightarrow Act(x) \wedge ACT(y)$
(Ap9) $Act(x) \rightarrow \exists! y(CO(x, y))$
(Ap10) $ACT(x) \rightarrow \exists y(CO(y, x))$
(Ap11) $ACT(x) \rightarrow \exists! y(AG(y) \wedge \forall z(CO(z, x) \rightarrow AGO(y, z))$
(Tp1) $Act(x) \rightarrow \exists! y(AG(y) \wedge AGO(y, x))$
(Dp4) $BAct(x, y) \triangleq \exists s(RT(y, s) \wedge BEG(x, s))$
(Dp5) $EAct(x, y) \triangleq \exists s(RT(y, s) \wedge END(x, s))$
(Dp6) $BACT(x, y) \triangleq \exists s(CO(s, y) \wedge BAct(x, s))$
(Ap12) $CO(x, z) \wedge CO(y, z) \rightarrow \exists t(BAct(t, x) \wedge BAct(t, y))$
(Dp7) $LON(x, y) \triangleq \exists s(RT(x, s) \wedge TP(s, y))$

3.3. Agency in OntoSTIT+

Understanding PO. To bind the intuitions that are behind *Choice*/PO within the OntoSTIT+ framework, we propose the formula (Ap13):

(Ap13) $CO(x, y) \wedge CO(z, y) \wedge AGO(u, x) \wedge BACT(w, y) \wedge LON(x, k) \wedge LON(z, l) \rightarrow PO(u, w, k, l)$,

which says that if x and z are action courses of an action token y with beginning w and agent u, then $PO(u, w, k, l)$ is underlying choice for action token y.

Filling the agentive gap. As we have just mentioned in section 2.2.4 actions themselves are not present in OntoSTIT and we were not able to express in it that *the agent switched off the light* by explicit referring to *switching* as such. In OntoSTIT+ we can easily do it.[9] Let's represent again the example (Es1):

(Ep1) $\exists x, z(ACT(x) \wedge$ *switching-off-the-light*$(x) \wedge \forall y, h(CO(y, x) \wedge LON(y, h) \rightarrow AGO(Robert, y) \wedge BAct(z, y) \wedge EAct(z, y) \wedge HOLDS(z, h,$*light-is-off*$)))$

Now, if we (i) loosen the condition that the action is instantaneous, i.e., that the beginning and end of each action course of a specific action token are the same, and (ii) limit the requirement that the action has been successful to the actual history **h** only, we obtain a description that captures also situations like that of example (Es4):

(Ep2) $\exists x(ACT(x) \wedge$ *killing-Lincoln*$(x) \wedge \forall y, z(CO(y, x) \wedge LON(y, \mathbf{h}) \wedge EAct(z, y) \rightarrow AGO(Booth, y) \wedge HOLDS(z, \mathbf{h},$*Lincoln is dead*$)))$

Notice that (Ep2) does not share the problems of (Es4) because the outcome of the action is linked to the action of the agent. By extending OntoSTIT on actions and intervals we solved two problems pointed out at the end of section 2.2.4.

[9] Here we are assuming the existence of a number of additional predicates, like *switching-off-the-light* and *killing-Lincoln*, that categorize action tokens.

4. Perspectives

In this work, we have proposed a first-order theory, OntoSTIT, that made explicit the ontological assumptions of the most expressive modal logic of agency to date, STIT. We have then showed how this framework could be extended, including actions in the domain of OntoSTIT+, to overcome some of STIT's shortcomings.

This is only a first step towards a rich theory of actions and agency. Obviously, OntoSTIT+ still needs to be extended in many directions. To deal with expected effects, which might be useful for, e.g., defining action categories, we can perhaps take inspiration from [17], specifying default actions courses. To deal more explicitly with the agent's intentions than with the simple 'possible outcomes' predicate, integrating agent's mental attitudes is a necessity. We also need to investigate how to express that different categories of actions unfold in time in different ways (aktionsart), and introduce other participants than the agent.

Before adding too many extensions, it might be interesting to take advantage of our departing point, a decidable propositional modal logic. We would thus like to study what is the decidable part of OntoSTIT+ and the possibility to transform it back into some modal logic extending directly STIT. Finally, the integration of OntoSTIT+ within a foundational ontology like DOLCE would surely bring many further insights.

References

[1] N. Troquard and L. Vieu. Towards a logic of agency and actions with duration. In *ECAI06*, 2006.

[2] J. F. Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.

[3] N. Belnap, M. Perloff, and M. Xu. *Facing the future: Agents and choices in our indeterminist world*. Oxford University Press, 2001.

[4] J. Broersen, A. Herzig, and N. Troquard. From Coalition Logic to STIT. In Wiebe van der Hoek, Alessio Lomuscio, Erik de Vink, and Mike Wooldridge, editors, *Third International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS 2005)*, volume 157(4) of *Electronic Notes in Theoretical Computer Science*, pp 23–35. Elsevier, 2006.

[5] J. Broersen, A. Herzig, and N. Troquard. Embedding Alternating-time Temporal Logic in Strategic STIT Logic of Agency, 2006. Submitted.

[6] E. M. Clarke and E. A. Emerson. Synthesis of synchronization skeletons for branching time temporal logic. *Logics of Programs Workshop*, volume131 of Lecture Notes in Computer Science, pages 52–71. Springer Verlag, 1981.

[7] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. In *Proceedings of the 38th Symposium on Foundations of Computer Science*, pp. 100–109. IEEE Computer Society Press, 1997.

[8] M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1): 149 –166, 2002.

[9] V. Goranko. Coalition games and alternating temporal logics. In *Proceedings of the 8th conference on Theoretical Aspects of Rationality and Knowledge*, pages 259–272. Morgan Kaufmann, 2001.

[10] D. Davidson. *Essays on Actions and Events*. Clarendon Press, Oxford, 1991.

[11] A. Goldman. *A Theory of Human Action*. Prentice-Hall, Englewood Cliffs, N. J., 1970.

[12] J. J. Thomson. *Acts and Other Events*. Cornell University Press, Ithaca, N.Y., 1977.

[13] M. Manzano. *Extentions of First Order Logic*. Cambridge University Press, 1996.

[14] H. J. Ohlbach. Combining Hilbert style and semantic reasoning in a resolution framework. *CADE-15, LNAI 1421*, pages 205–219, 1998.

[15] J. R. Searle. *Rationality in Action*. MIT Press, Cambridge Mass., 2001.

[16] P. M. Pietroski. *Causing Actions*. Oxford University Press, 2000.

[17] T. Müller. On the formal structure of continuous on the formal structure of continuous action. *Advances in Modal Logic*, 5:191–209, 2005.