A Computationally Grounded Logic of 'Seeing-to-it-that'

Andreas Herzig[1] , **Emiliano Lorini**[1] and **Elise Perrotin**[2]

[1]IRIT, CNRS, Toulouse, France
[2]CRIL, CNRS, Lens, France

{Andreas.Herzig,Emiliano.Lorini}@irit.fr, elise.perrotin@cril.fr

Abstract

We introduce a simple model of agency that is based on the concepts of control and attempt. Both relate agents and propositional variables. Moreover, they can be nested: an agent i may control whether another agent j controls a propositional variable p; i may control whether j attempts to change p; i may attempt to change whether j controls p; i may attempt to change whether j attempts to change p; and so on. In this framework we define several modal operators of time and agency: the LTL operators on the one hand, and the Chellas and the deliberative stit operator on the other. While in the standard stit framework the model checking problem is unfeasible because its models are infinite, in our framework models are represented in a finite and compact way: they are grounded on the primitive concepts of control and attempt. This makes model checking practically feasible. We prove its PSPACE-completeness and we show how the concept of social influence can be captured.

1 Introduction

One of the most prominent accounts of agency is Belnap and Horty's logical analysis of 'agent i sees to it that proposition φ is true in terms of 'branching time + agent choice' models (BT+AC models) [Horty and Belnap, 1995; Horty, 2001]. At the core of this so-called 'stit logic' are the Chellas stit operator, expressing the ability of a group of agents to ensure that a given formula is true regardless of the choices of other agents, and the deliberative stit operator, adding that truth of this formula is not necessary. This framework is very expressive and was put to work to account for other action-based concepts such as obligation [Horty, 2001; Broersen, 2011], influence [Lorini and Sartor, 2016], social commitment [Lorini, 2013] and responsibility [Lorini *et al.*, 2014]. Unfortunately reasoning turned out to be hard: deciding satisfiability of formulas involving the Chellas or the deliberative stit operator is NEXPTIME-complete even without temporal modalities as soon as there is more than one agent [Balbiani *et al.*, 2008]; it becomes 2EXPTIME-complete with temporal modalities 'next' and 'until' [Boudou and Lorini, 2018]; and it is undecidable as soon as agency of sets of

agents (groups) comes into play, and so again already without temporal modalities [Herzig and Schwarzentruber, 2008]. On the other hand, model checking—which is often considered to be an interesting alternative to satisfiability checking [Halpern and Vardi, 1991]—is basically unfeasible because BT+AC models are typically infinite.

We here propose to encode BT+AC models in a finite and compact way, so that model checking becomes practically feasible. We do so by grounding the basic concepts of stit logics, namely histories and choices, on the concepts of control and attempt. The former is borrowed from [van der Hoek *et al.*, 2011; Herzig *et al.*, 2011] while the latter is borrowed from [Lorini and Herzig, 2008]. Both relate agents and propositional variables: when an agent i controls a propositional variable p then she is able to determine the truth value of p at the next state; if this is the case and i attempts to change p then the truth value of p gets flipped at the next state; and if nobody is able and attempting to change the truth value of p then it remains unchanged.

We follow van der Hoek [2011] whose states are not just valuations of classical propositional logic (sets of propositional variables), but come equipped with a function associating to each propositional variable the agent controlling it. We basically augment these models by a function associating to each propositional variable the set of agents attempting to change it. States are therefore triples, and each triple determines a unique next state: all those variables whose change is attempted by some agent controlling it get their truth value flipped, and the other variables keep their truth value. Attempts are naturally viewed as persistent goals: agents abandon the attempt to change p once p has been successfully changed (possibly by somebody else).

Representing control and attempts in this way only allows for reasoning about the next state: as control does not change, attempts that fail at the first step will always fail, and new attempts cannot appear from one state to the next. We go beyond this simple model and introduce *higher-order control and attempt*: for two (possibly identical) agents i and j: i may control whether j controls a propositional variable p; i may control whether j attempts to change p; i may attempt to change whether j controls p; and i may attempt to change whether j attempts to change p. This allows us to reason about future states beyond the next state: we can define meaningful temporal operators of Linear-time Temporal

Logic (LTL). In these models we interpret modal operators of agency of the kind "group J achieves φ": we define a Chellas stit operator and a deliberative stit operator. For the resulting logic we establish PSPACE completeness of model checking.

The paper is organised as follows. In Section 2 we introduce the models and in Section 3 language and semantics of our logic of agency based on control and attempt (LACA). In Section 4 we show that they correspond to particular BT+AC models. Section 5 establishes complexity of model checking. Section 6 illustrates LACA by recasting the definition of social influence of [Lorini and Sartor, 2016]. Section 7 concludes.

2 Models of Control and Attempt

We define the models of the logic of agency based on control and attempt LACA. Let $\mathsf{Agt} = \{1, \ldots, \mathsf{card}(\mathsf{Agt})\}$ be a finite set of agent names, typically noted i, j,... *Groups* are subsets of Agt, typically noted J, J',... The complement of J is $\overline{J} = \mathsf{Agt} \setminus J$. We write $\bar{i}$ instead of $\overline{\{i\}}$.

2.1 Atoms and Valuations

Let Prp be a set of propositional variables, or *facts*, typically noted p, q,... Atomic formulas, or *atoms* for short, follow the grammar

$$\alpha ::= p \mid \mathsf{c}_i\alpha \mid \mathsf{t}_i\alpha,$$

where p ranges over Prp and i over Agt. $\mathsf{c}_i\alpha$ reads "agent i controls α" and $\mathsf{t}_i\alpha$ reads "agent i attempts to act on α", or "agent i attempts to change the truth value of α". Atoms are therefore sequences of c_i and t_i that are followed by a propositional variable: the variable the atom is *about*. Atoms of the form $\mathsf{c}_i\alpha$ are *control atoms* and atoms of the form $\mathsf{t}_i\alpha$ are *attempt atoms*. The set of all atoms is noted Atm.

States are simply *valuations over atoms*: subsets of the set Atm of atoms, noted V, V', etc. If $\alpha \in V$ then α is true at V, and if $\alpha \notin V$ then α is false at V. When $\mathsf{t}_i\alpha$ is false then i is indifferent about the value of p (as opposed to i wanting to maintain the truth value of p). We say that V and V' *agree* on α if α has the same status in V and V', that is, either $\alpha \in V$ and $\alpha \in V'$, or $\alpha \notin V$ and $\alpha \notin V'$. Otherwise V and V' *disagree* on α. Given a valuation V and a group of agents $J \subseteq \mathsf{Agt}$, $\mathsf{Att}_J(V) = \{\mathsf{t}_i\alpha \ : \ i \in J\}$ is the set of J's attempt atoms in V and $\mathsf{Ctrl}_J(V) = \{\mathsf{c}_i\alpha \ : \ i \in J\}$ is the set of J's control atoms of V, alias J's abilities at V.

Example 1. *Two agents 1 and 2 are in a room with a window and a radiator. The state of the room is described by the propositional variables w ('the window is closed') and h ('the heating is on'). 1 controls h and 2 controls w. Moreover, 1 has authority over 2 and can order 2 to open or close the window: 1 controls $\mathsf{t}_2 w$. Initially w and h are both false, and 1 is going to try to switch the heating on and to make 2 flip his goal about the window. This is modelled by the valuation*

$$V^{hw} = \{\mathsf{c}_1 h, \mathsf{t}_1 h, \mathsf{c}_2 w, \mathsf{c}_1 \mathsf{t}_2 w, \mathsf{t}_1 \mathsf{t}_2 w\}.$$

In our example control is exclusive: no variable is controlled by more than one agent. For the sake of generality we do not impose this as a general constraint.

2.2 Successor Valuations

A given valuation determines its successor valuation; let us discuss in detail how this should work.

We have said that $\mathsf{t}_i\alpha$ means that i is going to try to act on α. If i does not control α then this is going to fail; and if i controls α then the effect of i's trying is that the truth value of α gets flipped. For example, the successor of $V_1^h = \{\mathsf{c}_1 h, \mathsf{t}_1 h\}$ should contain h and the respective successors of $V_2^h = \{\mathsf{c}_1 h\}$ and $V_3^h = \{\mathsf{t}_1 h\}$ should not.

We understand an attempt atom $\mathsf{t}_i\alpha$ as a persistent goal in the sense of Cohen and Levesque's logic of intentions [1990] that is only abandoned once it has been satisfied: i will basically keep on trying until either the truth value of α has changed, or until an agent j (possibly i herself) terminates i's commitment to achieve α. When i has the goal to change the truth value of p then i typically also has the goal to abandon the goal $\mathsf{t}_i p$, i.e., to change the truth value of $\mathsf{t}_i p$. Therefore the valuations $V_1^h \cup \{\mathsf{c}_1 \mathsf{t}_1 h\}$ and $V_1^h \cup \{\mathsf{c}_1 \mathsf{t}_1 h, \mathsf{t}_1 \mathsf{t}_1 h\}$ should have the same successor, namely $\{h, \mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h\}$.

We consider that abilities (i.e., control) are *a priori* preserved, unless they get flipped by a successful higher-order attempt. Therefore the successor of the above V_1^h should intuitively be $\{\mathsf{c}_1 h, h\}$ and the successor of V_2^h should be V_2^h. We also consider that attempts should be preserved if they failed. Hence the successor of V_3^h should be V_3^h.

In our example it is possible that 1 achieves h (and therefore abandons her goal to flip h) and simultaneously makes 2 adopt the goal to flip h. Consider $V_1^h \cup \{\mathsf{c}_1 \mathsf{t}_2 h, \mathsf{t}_1 \mathsf{t}_2 h, \mathsf{c}_2 h\}$. On the one hand, 1's successful attempt to act on h will make 1 abandon her goal to act on h, i.e., $\mathsf{t}_1 h$ will become false; on the other hand, her successful second-order attempt to act on $\mathsf{t}_2 h$ will make the latter true. Hence the successor should be $\{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_2 h, \mathsf{c}_2 h, \mathsf{t}_2 h, h\}$, and the successor of the latter should be $\{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_2 h, \mathsf{c}_2 h\}$. Overall, h was initially false, then became true, and ends up false again.

The *changeset* of a valuation V is

$$V^{\pm} = \{\alpha \ : \ \mathsf{c}_i\alpha \in V \text{ and } \mathsf{t}_i\alpha \in V \text{ for some } i \in \mathsf{Agt}\}.$$

The elements of $V^{\pm}$ are the goals that should a priori be abandoned: the atoms of V whose change of truth value is directly caused by the agents' successful attempts in V. The definition of the successor of a valuation then is:

$$succ(V) = (V^{\pm} \setminus V) \cup$$
$$\left(V \setminus \left(V^{\pm} \cup \{\mathsf{t}_i\alpha \in V \ : \ \alpha \in V^{\pm}, i \in \mathsf{Agt}\}\right)\right).$$

Hence (1) everything in the changeset that is false becomes true; (2) everything that is true and is neither in the changeset nor a goal to be abandoned remains true. Put differently: V and $succ(V)$ agree on α if and only if $\alpha \notin V^{\pm}$ and there is no $\beta \in V^{\pm}$ such that $\alpha = \mathsf{t}_i\beta$. Some examples are:

$$succ(\{\mathsf{c}_1 h, \mathsf{c}_2 h, \mathsf{t}_1 h\}) = \{\mathsf{c}_1 h, \mathsf{c}_2 h, h\},$$
$$succ(\{\mathsf{c}_1 h, \mathsf{c}_2 h, \mathsf{t}_1 h, \mathsf{t}_2 h\}) = \{\mathsf{c}_1 h, \mathsf{c}_2 h, h\},$$
$$succ(\{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, h\}) = \{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, h\},$$
$$succ(\{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, \mathsf{t}_1 \mathsf{t}_1 h\}) = \{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, \mathsf{t}_1 h\},$$
$$succ(\{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, \mathsf{t}_1 \mathsf{t}_1 h, \mathsf{t}_1 h\}) = \{\mathsf{c}_1 h, \mathsf{c}_1 \mathsf{t}_1 h, h\}.$$

The function $succ(.)$ is total: every valuation has exactly one successor. A valuation may have more than one predecessor, as illustrated by the first two lines of our examples. (Another example is $succ(\emptyset) = succ(\mathsf{Atm}) = \emptyset$.)

We inductively define the *n-th successor* of V:

$$succ^0(V) = V,$$
$$succ^{n+1}(V) = succ(succ^n(V)).$$

Example 2 (Example 1, ctd.). *In the valuation*

$$V^{hw} = \{\mathsf{c}_1 h, \mathsf{t}_1 h, \mathsf{c}_2 w, \mathsf{c}_1 \mathsf{t}_2 w, \mathsf{t}_1 \mathsf{t}_2 w\},$$

agent 1 *is going to try to switch the heating on and to make* 2 *flip his goal about the window. Both attempts are going to be successful because* 1 *controls* h *and* 1 *controls* 2*'s attempts on* w. *Hence at* $succ(V)$ *the heating is on and the window is open, and at* $succ^2(V)$ *and beyond the heating is on and the window is closed:*

$$succ^1(V) = \{\mathsf{c}_1 h, h, \mathsf{c}_2 w, \mathsf{c}_1 \mathsf{t}_2 w, \mathsf{t}_2 w\},$$
$$succ^n(V) = \{\mathsf{c}_1 h, h, \mathsf{c}_2 w, \mathsf{c}_1 \mathsf{t}_2 w, w\}, \ \textit{for } n \geqslant 2.$$

2.3 Agent Choices

We define the relation $\sim_J$ between valuations as follows:

$V \sim_J V'$ if V and V' agree on every atom in $\mathsf{Prp} \cup \mathsf{Ctrl} \cup$
$$\{\mathsf{t}_i\alpha \ : \ i \in J, \alpha \in \mathsf{Prp} \cup \mathsf{Ctrl} \text{ and } \mathsf{c}_i\alpha \in V\}.$$

That is, $V \sim_J V'$ iff V and V' agree on facts, abilities, and on successful attempts by agents in J to flip facts and abilities at the next step. We call the latter the *immediate choices* of J at V. Put differently, $V \sim_J V'$ when V and V' agree on facts and abilities as well as in the way that agents of J collectively influence the evolution of these facts and abilities in the immediate future. For example, for the valuation of Example 2 we have $V^{hw} \sim_{\{1\}} \{\mathsf{c}_1 h, \mathsf{t}_1 h, \mathsf{c}_2 w, \mathsf{c}_1 \mathsf{t}_2 w, \mathsf{t}_2 w\}$. When $V \sim_\emptyset V'$ then V and V' agree on facts and abilities, abbreviated $V \sim V'$. The relations $\sim_J$ are reflexive, transitive, and symmetric. Moreover, $V \sim_{\mathsf{Agt}} V'$ implies $succ(V) \sim succ(V')$.

We generalize this and define the relation $\sim_J^n$ between valuations as follows, for $n \in \mathbb{N}_0$:

$V \sim_J^n V'$ if $succ^n(V) \sim_J succ^n(V')$ and
$$\text{for every } m < n, succ^m(V) \sim_{\mathsf{Agt}} succ^m(V').$$

That is, $V \sim_J^n V'$ if V and V' agree on facts and abilities and on immediate choices of all agents up to step $n-1$; and moreover agree on immediate choices of J at step n.

Hence $V \sim_J^0 V'$ iff $V \sim_J V'$. We define $\sim^n = \sim_\emptyset^n$. That is, $V \sim^n V'$ if $succ^m(V) \sim_{\mathsf{Agt}} succ^m(V')$ for all $m < n$.

3 The Logic LACA

We now define language and semantics of LACA.

3.1 Language of LACA

The language of LACA, $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Atm})$, is defined by the following grammar:

$$\varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathtt{X}\varphi \mid \mathtt{G}\varphi \mid [J \operatorname{cstit}]\varphi \mid [J \operatorname{dstit}]\varphi,$$

where α ranges over Atm and J ranges over the powerset of Agt. The formula $\mathtt{X}\varphi$ reads "φ holds next"; $\mathtt{G}\varphi$ reads "φ holds henceforth"; $[J \operatorname{cstit}]\varphi$ reads "J sees to it that φ" (the so-called Chellas stit operator); and $[J \operatorname{dstit}]\varphi$ reads "J deliberately sees to it that φ" (the deliberative stit operator). The language $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$ is the fragment of $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Atm})$ without control and attempt atoms and is the language of the original group stit logic (in its discrete time version).

When convenient we drop set parentheses of coalitions and write e.g. $[1, 2 \operatorname{cstit}]p$ instead of $[\{1, 2\} \operatorname{cstit}]p$. Historic necessity is defined as $\Box\varphi = [\emptyset \operatorname{cstit}]\varphi$. The dual of historic necessity is $\Diamond\varphi = \neg\Box\neg\varphi$ and that of the cstit operator is $\langle J \operatorname{cstit}\rangle\varphi = \neg[J \operatorname{cstit}]\neg\varphi$.

We define the set of atoms of a formula φ as follows:

$$\mathsf{Atm}(\alpha) = \{\alpha\},$$
$$\mathsf{Atm}(\varphi \wedge \psi) = \mathsf{Atm}(\varphi) \cup \mathsf{Atm}(\psi),$$
$$\mathsf{Atm}(*\varphi) = \mathsf{Atm}(\varphi) \text{ for } * \in \{\neg, \mathtt{X}, \mathtt{G}, [J \operatorname{cstit}], [J \operatorname{dstit}]\}.$$

For example, $\mathsf{Atm}(\mathsf{c}_1\mathsf{t}_2 p \wedge [1, 2 \operatorname{cstit}]\mathsf{c}_3 p) = \{\mathsf{c}_1\mathsf{t}_2 p, \mathsf{c}_3 p\}$.

3.2 Truth Conditions

Here are the truth conditions for $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Atm})$ formulas:

$$
\begin{aligned}
&V, n \models \alpha && \text{if } \alpha \in succ^n(V), \\
&V, n \models \mathtt{X}\varphi && \text{if } V, n+1 \models \varphi, \\
&V, n \models \mathtt{G}\varphi && \text{if } V, m \models \varphi \text{ for every } m \geqslant n, \\
&V, n \models [J \operatorname{cstit}]\varphi && \text{if } V', n \models \varphi \text{ for every } V' \sim_J^n V, \\
&V, n \models [J \operatorname{dstit}]\varphi && \text{if } V, n \models [J \operatorname{cstit}]\varphi \text{ and } V, n \not\models \Box\varphi,
\end{aligned}
$$

and as usual for the Boolean operators. The idea underlying the Chellas stit modality $[J \operatorname{cstit}]\varphi$ is that the current choices of the agents in J guarantee that φ is true no matter what the opponents in $\overline{J}$ attempt. This is the same as: it is not the case that the agents in $\overline{J}$ are able to achieve $\neg\varphi$ while the agents in J stick to what they currently attempt. We can therefore view $[J \operatorname{cstit}]$ as an operator of inability of J's opponents to avoid φ. The deliberative stit modality adds to the Chellas stit a negative condition: for J to deliberately see to it that φ, φ should not be necessarily true.

Example 3 (Example 2, ctd.). *We have*

$$V^{hw}, 0 \models \neg h \wedge \neg w \wedge \mathtt{X}(h \wedge \neg w) \wedge \mathtt{X}\mathtt{X}(h \wedge w) \wedge$$
$$[1 \operatorname{cstit}]\mathtt{X}h \wedge \mathtt{X}[2 \operatorname{cstit}]\mathtt{X}w.$$

The formula remains true when the cstit operator is replaced by the dstit operator.

A $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Atm})$ formula φ is satisfiable if $V, 0 \models \varphi$ for some V; and it is valid if $V, 0 \models \varphi$ for every V.

Example 4. *In Homer's Odyssey, Ulysses escapes the sirens by removing his own control to follow their calls. This example is considered by the philosopher Jon Elster to be paradigmatic of the idea that self-control can be used for coping with weakness of will [Elster, 1979]. We model two versions of the situation: what might have happened if Ulysses had not been so clever to get tied to the mast and what actually happens. Initially Ulysses (u) is not on the sirens' island and controls whether he is on the island or not ($\mathsf{c}_u \mathtt{OnIs}$). He can abandon his control ($\mathsf{c}_u \mathsf{c}_u \mathtt{OnIs}$). The sirens, for simplicity viewed as a single agent s, can make Ulysses change his mind about*

swimming to the island (`c_s t_u OnIs`) and indeed attempt to do so (`t_s t_u OnIs`). This corresponds to the following valuation:

$$V_1^u = \{ \mathtt{c_u OnIs}, \mathtt{c_u c_u OnIs}, \mathtt{c_s t_u OnIs}, \mathtt{t_s t_u OnIs} \}.$$

But Ulysses actually decides to abandon his control by getting tied to the mast: the valuation corresponding to Homer's story is $V_2^u = V_1^u \cup \{ \mathtt{t_u c_u OnIs} \}$. Then $V_1^u \sim_s V_2^u$ and:

$$succ(V_1^u) = \{ \mathtt{c_u OnIs}, \mathtt{c_u c_u OnIs}, \mathtt{c_s t_u OnIs}, \mathtt{t_u OnIs} \},$$
$$succ(V_2^u) = \{ \mathtt{c_u c_u OnIs}, \mathtt{c_s t_u OnIs}, \mathtt{t_u OnIs} \}.$$

Therefore:

$$V_1^u, 0 \models \mathtt{X}[s\,\mathrm{cstit}]\mathtt{XOnIs},$$
$$V_2^u, 0 \models [u\,\mathrm{dstit}]\mathtt{XX}\neg\mathtt{OnIs} \wedge [u\,\mathrm{dstit}]\mathtt{X}[u\,\mathrm{cstit}]\mathtt{X}\neg\mathtt{OnIs}.$$

In V_1^u, the sirens get their wishes and see to it that Ulysses follow their calls. In V_2^u, Ulysses manages to avoid this outcome and (deliberately) sees to it that he stays on his ship by choosing to remove his own control.

4 Reconstruction of BT+AC Structures

We now assess our control- and attempt-based logic w.r.t. Belnap and Horty's stit logic by establishing the relation with their semantics in terms of BT+AC models. Our construction leads to a discrete time structure and inverses the standard way the definition of BT+AC models proceeds. The latter takes moments that are ordered in time to be primitives, then defines histories, and finally adds agent choices. In contrast, our construction starts from histories (which are induced by valuations via the successor function) and then defines moments, their temporal ordering, and agent choices. We start by recalling BT+AC models.

4.1 Discrete BT+AC Models

We recall the definitions from [Horty, 2001], which we adapt to the case of discrete tree structures as done e.g. in [Boudou and Lorini, 2018; Broersen *et al.*, 2006; Schwarzentruber, 2012]. A *discrete BT structure* is a pair $(W, <)$ where W is a non-empty set of moments and $<$ is a strict partial order on W that is tree-like[1] and discrete. *Histories* are maximal sets of linearly ordered moments. The set of all histories of $(W, <)$ is *Hist*. The set of histories sharing the moment w is $H_w \subseteq Hist$. The (unique) successor of w on history h is $succ(w/h)$. A BT+AC model is a quadruple $\mathcal{M} = (W, <, Choice, \mathcal{V})$ where $(W, <)$ is a BT structure, *Choice* maps every moment-agent pair (w, i) to a partition of H_w, and $\mathcal{V}$ associates to each $p \in \mathsf{Prp}$ a set of moment-history pairs: $\mathcal{V}(p) \subseteq W \times Hist$. Essentially, H_w represents all possible timelines branching from moment w, each corresponding to different choices of agents at w, and $Choice(w, i)$ represents the possible choices of a single agent i at w, each leading to a different subset of H_w. It is assumed that *Choice* is such that:[2]

1. For every $Q \in Choice(w, i)$ we have $Q \neq \emptyset$.

---

2. For every w and mapping $s_w : \mathsf{Agt} \longrightarrow 2^{H_w}$ such that $s_w(i) \in Choice(w, i)$ (that is, a selection of one choice for each agent) we have $\bigcap_{i \in \mathsf{Agt}} s_w(i) \neq \emptyset$.

3. If h and h' are undivided at w then $h' \in Choice(w, i)(h)$.[3]

The second constraint says that the agents' choices are independent. The third constraint is called 'no choice between undivided histories'. Letting the set of all selection functions at w be

$$Select_w = \{ s_w : s_w(i) \in Choice(w, i),\ \text{for every } i \in \mathsf{Agt} \},$$

choices of groups of agents J are defined by:

$$Choice(w, J) = \Big\{ \bigcap_{i \in J} s_w(i) : s_w \in Select_w \Big\}.$$

In particular, $Choice(w, \emptyset) = \{H_w\}$ for the empty group. Moreover, given a history $h \in H_w$,

$$Choice(w, J)(h) = \{ h' \in H_w : h, h' \in Q$$
$$\text{for some } Q \in Choice(w, J) \}.$$

That is, $Choice(w, J)(h)$ is the particular set of histories containing h among the choices of group J, or in other words, the choice of group J at moment w in the history h.

Then the formulas of the language given in Section 3.1 are interpreted at moment-history pairs as follows:

$\mathcal{M}, w/h \models p$ if $w/h \in \mathcal{V}(p)$,
$\mathcal{M}, w/h \models \mathtt{X}\varphi$ if $\mathcal{M}, succ(w/h)/h \models \varphi$,
$\mathcal{M}, w/h \models [J\,\mathrm{cstit}]\varphi$ if $\mathcal{M}, w/h' \models \varphi$ for every
                                  $h' \in Choice(w, J)(h)$,
$\mathcal{M}, w/h \models [J\,\mathrm{dstit}]\varphi$ if $\mathcal{M}, w/h \models [J\,\mathrm{cstit}]\varphi$ and
                                  $\mathcal{M}, w/h \not\models \Box\varphi$,

and as expected for the Boolean operators.

In the rest of the section we show how to construct BT+AC structures from the valuations of Section 2. We start by defining histories as mappings from natural numbers to valuations and then show how every valuation generates a history.

4.2 From Valuations to BT Structures

In the context of our models of control and attempt, moments can be identified with $\sim^n$ equivalence classes, choices correspond to agents' attempts, and branching histories stem from different choices for agents.

We start with BT structures. Let V_0 be a valuation. Let $V_0^\sim$ be the $\sim$-equivalence class of V_0: $V_0^\sim = \{ V : V \sim V_0 \}$.

The set of moments is defined as

$$W_{V_0} = \biguplus_{n \in \mathbb{N}_0} W_{V_0}^n,$$

where for all $n \in \mathbb{N}_0$, $W_{V_0}^n = V_0^\sim / \sim^n$ is the quotient operator. We use disjoint union ($\uplus$) in order to ensure that two equivalent moments, whereby agents control the same variables and make the same attempts, can occur at different depths of the BT structure. Given a moment $w \in W_{V_0}^n$, the set of successors of w for the order $<_{V_0}$ is:

$$\mathbf{succ}(w) = \{ w' \in W_{V_0}^{n+1} : w \cap w' \neq \emptyset \}.$$

---

[1] If $w_1 < w_3$ and $w_2 < w_3$ then $w_1 = w_2$ or $w_1 < w_2$ or $w_2 < w_1$.

[2] It is often assumed that $Choice(w, i) \neq \emptyset$, but this is redundant.

[3] Two histories h and h' are *undivided* at a moment $w \in W$ if either w and $succ(w/h)$ are both on h', or w has no successor. That is, if w has a successor then w has the same successor in h and h'.

Fact 1. $\text{succ}(w)$ *is a partition of* w.

Proposition 1. $<_{V_0}$ *is a tree-like order.*

Recall that histories in BT structures are maximal sets of linearly ordered moments. If h is a history of $(W_{V_0}, <_{V_0})$, then for all $n \in \mathbb{N}_0$ we denote by $h(n)$ the unique member of $h \cap W_{V_0}^n$. Every valuation V clearly induces a single history h_V such that $V \in h_V(n)$ for all $n \in \mathbb{N}_0$. The other way round, each history is induced by at least one valuation.

Proposition 2. *For every history h of $(W_{V_0}, <_{V_0})$, there exists $V \sim V_0$ such that $h = h_V$.*

Corollary 1. *For every $w \in W_{V_0}$, the set H_w of histories containing w is exactly the set of histories induced by valuations in w, i.e., we have $H_w = \{h_V \ : \ V \in w\}$.*

4.3 From Valuations to Agent Choices

We now show how to construct a full BT+AC model $\mathcal{M}_{V_0} = (W_{V_0}, <_{V_0}, Choice_{V_0}, \mathcal{V}_{V_0})$ from a valuation V_0. We have already defined W_{V_0} and $<_{V_0}$. As all valuations in an equivalence class for $\sim$ agree on propositional variables, $\mathcal{V}_{V_0}$ is naturally defined by

$$\mathcal{V}_{V_0}(p) = \{w/h \ : \ w \in W_{V_0}^n \text{ for some } n \in \mathbb{N}_0, h \in H_w,$$
$$\text{and } p \in succ^n(V) \text{ for } V \in w\}.$$

If i is an agent and $w \in W_{V_0}^n$, we define

$$\mathsf{Ctrl}_i(w) = \{\alpha \in \mathsf{Prp} \cup \mathsf{Ctrl} \ : \ \text{there is a } V \in w \text{ such that}$$
$$\mathsf{c}_i\alpha \in succ^n(V) \text{ and } \mathsf{t}_i\alpha \in succ^n(V)\}.$$

Intuitively, the set $\mathsf{Ctrl}_i(w)$ is the set of possible immediate choices of agent i at w. We now define $Choice_{V_0}(w, i)$ for $w \in W_{V_0}^n$ as:

$$Choice_{V_0}(w, i) = \{\{h_V \ : \ V \in w \text{ and}$$
$$\mathsf{Att}_i(succ^n(V)) \cap (\mathsf{Prp} \cup \mathsf{Ctrl}) = \Gamma\} \ :$$
$$\Gamma \subseteq \mathsf{Ctrl}_i(w)\}.$$

Proposition 3. *Consider $w \in W_{V_0}$. Then:*

1. *$Choice_{V_0}(w, i)$ is a partition of H_w;*

2. *Any $H \in Choice_{V_0}(w, i)$ is nonempty;*

3. *$Choice_{V_0}(w, i)$ satisfies the "no choice between undivided histories" constraint;*

4. *For every selection function $s_w : \mathsf{Agt} \longrightarrow 2^{H_w}$ such that $s_w(i) \in Choice_{V_0}(w, i)$ we have $\bigcap_{i \in \mathsf{Agt}} s_w(i) \neq \emptyset$.*

The choices of group J at w are:

$$Choice_{V_0}(w, J) = \{\{h_V \ : \ V \in w \text{ and } \forall i \in J,$$
$$\mathsf{Att}_i(succ^n(V)) \cap (\mathsf{Prp} \cup \mathsf{Ctrl}) = \Gamma_i\} \ :$$
$$\forall i \in J, \Gamma_i \subseteq \mathsf{Ctrl}_i(w)\}.$$

Given a history h_V with $w = h_V(n)$, $Choice_{V_0}(w, J)(h_V)$ corresponds to fixing the immediate attempts of agents of J at $h_V(n)$ to those of $succ^n(V)$:

$$Choice_{V_0}(w, J)(h_V) = \{h_{V'} \ : \ V' \in w \text{ and}$$
$$succ^n(V') \sim_J succ^n(V)\}$$
$$= \{h_{V'} \ : \ V' \sim_J^n V\}.$$

Proposition 4. $\mathcal{M}_{V_0} = (W_{V_0}, <_{V_0}, Choice_{V_0}, \mathcal{V}_{V_0})$ *is a BT+AC structure satisfying all constraints of Section 4.1.*

4.4 Equivalence

We are ready to show that all standard stit validities are also valid in models of control and attempt.

Proposition 5. *Let $\varphi \in \mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$ be a formula and $V_0 \in 2^{\mathsf{Atm}}$ be a valuation. For all $n \in \mathbb{N}_0$ and $V \sim V_0$, $\mathcal{M}_{V_0}, h_V(n)/h_V \models \varphi$ iff $V, n \models \varphi$.*

4.5 Beyond Standard Validities

The above inclusion is strict: there are LACA-valid $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$ formulas that not valid in BT+AC models. First of all, standard validities in BT+AC models with discrete time such as $[J \, \mathrm{cstit}]\mathsf{X}\varphi \to \mathsf{X}\square\varphi$. are also LACA valid. Beyond that, facts and abilities are *moment determinate* in LACA, as expressed by the implications $\alpha \to \square\alpha$ and $\neg\alpha \to \square\neg\alpha$, for $\alpha \in \mathsf{Prp} \cup \mathsf{Ctrl}$. This property is discussed as optional in the stit literature [Horty, 2001].

Here are three formulas that are not standardly valid but are valid in LACA. The implication $\square\mathsf{X}p \to p$, for $p \in \mathsf{Prp}$ expresses that facts don't change by themselves, and $(\neg p \wedge \mathsf{X}p) \to \bigvee_{i \in \mathsf{Agt}}[i\,\mathrm{cstit}]\mathsf{X}p$ expresses that every fact change is caused by some agent. Finally, the equivalences $[i\,\mathrm{cstit}](p \vee q) \leftrightarrow ([i\,\mathrm{cstit}]p \vee [i\,\mathrm{cstit}]q)$ and $[J\,\mathrm{cstit}]p \leftrightarrow \bigvee_{i \in J}[i\,\mathrm{cstit}]p$ are valid because in LACA, agency is based on a relation between individuals and atoms.

5 Complexity of Model Checking

The model checking problem we are interested in is the following: given a valuation $V \subseteq \mathsf{Atm}$, a time point $n \in \mathbb{N}_0$ and formula $\varphi \in \mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$, do we have $V, n \models \varphi$? In this section we prove that it is PSPACE-complete.

5.1 Relevant Valuations

A key factor in the complexity of evaluating formulas is the infinity of valuations that need to be checked when evaluating G and $[J\,\mathrm{cstit}]$ operators. Let us investigate whether we can restrict ourselves to considering only a finite number of valuations. First, we define the *length* of an atom of Atm to be the length of the prefix of the propositional variable:

$$\ell(p) = 0, \quad \ell(\mathsf{c}_i\alpha) = 1 + \ell(\alpha), \quad \ell(\mathsf{t}_i\alpha) = 1 + \ell(\alpha).$$

For example, $\ell(\mathsf{c}_1\mathsf{c}_2\mathsf{t}_3p) = 3$. We extend the definition of length to finite valuations: $\ell(V) = \sum_{\alpha \in V} \ell(\alpha)$.

We can now establish that finite valuations only have a finite number of successors:

Proposition 6. *Let $V \subseteq \mathsf{Atm}$ be a finite valuation.*

1. *$\mathsf{card}(V) \geqslant \mathsf{card}(succ(V))$;*

2. *Either $\ell(V) > \ell(succ(V))$ or $succ(V) = V$.*

Given a valuation V and a group of agents J, considering all valuations such that $V \sim_J^n V'$ amounts to considering variations on $\overline{J}$'s attempts as well as on attempts of agents of J outside of their immediate choices. There is actually no need to consider variations on attempt atoms that are not a suffix of any control atom in V.

Proposition 7. *Let V be a valuation, J be a group of agents and φ be a formula of $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$. Then $V, n \models [J \mathrm{\,cstit}]\varphi$ iff $V', n \models \varphi$ for all V' such that $V' \sim_J^n V$ and V and V' differ only by attempt atoms $\mathtt{t}_i \alpha$ such that $\mathtt{c}_i \alpha$ is a suffix of a control atom in $\mathsf{Ctrl}_{\mathsf{Agt}}(V)$.*

5.2 PSPACE-Completeness

Using the results of the previous section, we now prove that model checking formulas of $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$ in our framework is PSPACE-complete.

Proposition 8. *The $\mathcal{L}_{\mathsf{LACA}}(\mathsf{Prp})$ model checking problem is in PSPACE-complete.*

Proof sketch. The membership proof is based on an algorithm computing whether $V, n \models \varphi$. The number and size of alternative valuations that need to be checked in the cases of $\mathsf{G}\varphi$, $[J \mathrm{\,cstit}]\varphi$ and $[J \mathrm{\,dstit}]\varphi$ are limited by Propositions 6 and 7.

Hardness is shown through a polynomial reduction of the problem of model checking Quantified Boolean Formulas (QBF), which is known to be PSPACE-complete. □

6 Application: Modelling Influence

In this section we we put our logic LACA into practice and show how it allows us to formalise Lorini and Sartor's [2016] definition of social influence. Their concept is based on the idea that for the influencer i to induce an influencee j to make a certain choice, i has to change j's *choice context*, so that a choice differing from j's preferred option without this modification becomes preferable to j. Agent j's choice context includes both j's available choices determined by what agent j controls (i's *choice set*) and j's preferences. "i influences j so that φ" is defined as:

$$[i \mathrm{\,infl\,} j]\varphi = [i \mathrm{\,cstit}]\mathtt{X}(\mathsf{Rat}_j \to [j \mathrm{\,dstit}]\varphi),$$

where Rat_j expresses that j is rational. For agent i (the influencer) to induce agent j (the influencee) to see to it that φ, i's actual choice must guarantee that every rational choice of j will make φ true. The reason why agent j's action is modeled by the deliberative stit operator is that for agent j to be influenced by agent i, j should have a choice available in his choice set possibly leading to $\neg\varphi$; i.o.w., influence requires that the consequence of the influencee's action could have been avoided, had the influencee made a different choice.

Let us see how we can express that an agent j is rational. Simply supposing that Rat_j is yet another special propositional variable (as done in [Lorini and Sartor, 2016]) fails to do the job because that variable would not be related to the agent's choices. Instead, we suppose that for every agent i there is a *formula* Rat_i characterising the states where an outcome with maximal utility obtains next: an agent is rational in a valuation if the successor valuation is ideal for the agent. One possibility is that Rat_i is of the form $\mathtt{X}\varphi$: i prefers φ to be true next; in particular, when i prefers φ to be true from the next state on then Rat_i contains $\mathtt{XG}\varphi$. Another possibility is that Rat_i characterises, for each action in i's choice set, under which conditions it is rational for i to try to perform it. Then Rat_i is typically a conjunction of equivalences of the

form $\mathtt{t}_i p \leftrightarrow \varphi_p^i$ where φ_p^i does not involve attempt atoms. For example, if i prefers to make p true whatever the values of the other variables are then Rat_i contains $\mathtt{t}_i p \leftrightarrow \neg p$.

Example 5 (Example 1, ctd.)**.** *Suppose agent 2 wants to to save energy: she prefers that the window is closed (w) when the heating is on (h); otherwise she does not care. We therefore have $\mathsf{Rat}_2 = \mathtt{X}(h \to w)$. Consider the valuation*

$$V_1^{hw} = \{\mathtt{c}_1 h, \mathtt{t}_1 h, \mathtt{c}_2 w, \mathtt{c}_2 \mathtt{t}_2 w\}.$$

Hence 1 is going to try to switch the heating on, which, intuitively, should influence 2 so that the window is going to be closed. In terms of actions: 1's switching the heating on is going to make 2 close the window. We therefore expect $V_1^{hw}, 0 \models [1 \mathrm{\,infl\,} 2]\mathtt{X}w$, i.e., $V_1^{hw}, 0 \models [1 \mathrm{\,cstit}]\mathtt{X}(\mathsf{Rat}_2 \to [2 \mathrm{\,dstit}]\mathtt{X}w)$. Let us check in detail that this is indeed the case. First, when interpreting the modality $[1 \mathrm{\,cstit}]$ at V_1^{hw} we check whether $V_1^{hw} \cup S, 0 \models \mathtt{X}(\mathsf{Rat}_2 \to [2 \mathrm{\,dstit}]\mathtt{X}w)$, for every $S \subseteq \{\mathtt{t}_2\alpha : \alpha \in \mathsf{Atm}\} \cup \{\mathtt{t}_1\alpha : \mathtt{c}_1\alpha \notin V \text{ or } \alpha = \mathtt{t}_i\beta\}$. Thanks to Proposition 7, we know that it is enough to consider variations on $\mathtt{t}_2 w$ and $\mathtt{t}_2\mathtt{t}_2 w$. Let us consider the three sets V_1^{hw}, $V_2 = V_1^{hw} \cup \{\mathtt{t}_2 w\}$, and $V_3 = V_1^{hw} \cup \{\mathtt{t}_2\mathtt{t}_2 w\}$. Their respective successors are

$$succ(V_1^{hw}) = \{h, \mathtt{c}_1 h, \mathtt{c}_2 w, \mathtt{c}_2 \mathtt{t}_2 w\},$$
$$succ(V_2) = \{h, \mathtt{c}_1 h, \mathtt{c}_2 w, \mathtt{c}_2 \mathtt{t}_2 w, w\},$$
$$succ(V_3) = \{h, \mathtt{c}_1 h, \mathtt{c}_2 w, \mathtt{c}_2 \mathtt{t}_2 w, \mathtt{t}_2 w\}.$$

Observe that $succ(V_2)$ is obtained from V_1^{hw} by 1 switching the heating on and 2 simultaneously closing the window, so to speak anticipating 1's influence. Both $(V_2, 1)$ and $(V_3, 1)$ satisfy Rat_2, and it therefore remains to check whether $V_2, 1 \models [2 \mathrm{\,dstit}]\mathtt{X}w$ and $V_3, 1 \models [2 \mathrm{\,dstit}]\mathtt{X}w$, which are indeed the case. Therefore $V_1^{hw}, 0 \models [1 \mathrm{\,infl\,} 2]\mathtt{X}w$.

7 Conclusion

We have presented a 'stit' framework in which the notion of action is conceptually and computationally grounded on the notions of control and attempt. We have proved that, unlike the standard 'stit' framework in which model checking is unfeasible because histories are infinite, model checking in our framework is PSPACE-complete. In future work, we expect to broaden its scope of application by formalizing further action-related concepts that have been investigated in the literature on 'stit' logic. The latter include the concept of controllability, as an agent i's capacity to influence another agent j to perform a certain action [Lorini and Sartor, 2021], as well as the concept of responsibility in its active form (i.e., responsibility for action) and passive form (i.e., responsibility for omission) [Lorini et al., 2014; Lorini and Schwarzentruber, 2011]. As we have shown in Section 6, our logic LACA offers a minimalistic and computionally grounded framework for modeling these concepts. Moreover, we plan to come up with an axiomatisation of the validities of our discrete group stit logic. Last but not least, we plan to enrich our temporal stit framework with past tense modalities for 'yesterday' (Y) and 'always in the past' (H) and to formalize the concept of 'achievement stit' studied in [Belnap et al., 2001] that better approximates agent causation.

Acknowledgments

References

[Balbiani *et al.*, 2008] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.

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

[Boudou and Lorini, 2018] Joseph Boudou and Emiliano Lorini. Concurrent game structures for temporal STIT logic. In Elisabeth André, Sven Koenig, Mehdi Dastani, and Gita Sukthankar, editors, *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, pages 381–389. IFAAMAS/ACM, 2018.

[Broersen *et al.*, 2006] Jan Broersen, Andreas Herzig, and Nicolas Troquard. Embedding Alternating-time Temporal Logic in strategic STIT logic of agency. *Journal of Logic and Computation*, 16(5):559–578, 2006.

[Broersen, 2011] Jan Broersen. Deontic epistemic stit logic distinguishing modes of mens rea. *Journal of Applied Logic*, 9(2):137–152, 2011.

[Cohen and Levesque, 1990] Philip R Cohen and Hector J Levesque. Intention is choice with commitment. *Journal of Artificial intelligence*, 42(2):213–261, 1990.

[Elster, 1979] Jon Elster. *Ulysses and the Sirens*. Cambridge University Press, Cambridge, 1979.

[Halpern and Vardi, 1991] Joseph Y. Halpern and Moshe Y. Vardi. Model checking vs. theorem proving: A manifesto. In James F. Allen, Richard Fikes, and Erik Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Cambridge, MA, USA, April 22-25, 1991*, pages 325–334. Morgan Kaufmann, 1991. Also in Artificial and Mathematical Theory of Computation 1991: 151-176.

[Herzig and Schwarzentruber, 2008] Andreas Herzig and Francois Schwarzentruber. Properties of logics of individual and group agency. In Carlos Areces and Robert Goldblatt, editors, *Advances in Modal Logic (AiML)*, pages 133–149, Nancy, 2008. College Publications.

[Herzig *et al.*, 2011] Andreas Herzig, Emiliano Lorini, Nicolas Troquard, and Frédéric Moisan. A dynamic logic of normative systems. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 228–233, 2011.

[Horty and Belnap, 1995] John Horty and Nuel Belnap. The deliberative stit: a study of action, omission, ability and obligation. *Journal of Philosophical Logic*, 24(6):583–644, 1995.

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

[Lorini and Herzig, 2008] Emiliano Lorini and Andreas Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.

[Lorini and Sartor, 2016] Emiliano Lorini and Giovanni Sartor. A STIT logic for reasoning about social influence. *Studia Logica*, 104(4):773–812, 2016.

[Lorini and Sartor, 2021] Emiliano Lorini and Giovanni Sartor. The ethics of controllability as influenceability. In Schweighofer Erich, editor, *Legal Knowledge and Information Systems - JURIX 2021: The Thirty-fourth Annual Conference, Vilnius, Lithuania, 8-10 December 2021*, volume 346 of *Frontiers in Artificial Intelligence and Applications*, pages 245–254. IOS Press, 2021.

[Lorini and Schwarzentruber, 2011] Emiliano Lorini and Francois Schwarzentruber. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 175(3-4):814–847, 2011.

[Lorini *et al.*, 2014] Emiliano Lorini, Dominique Longin, and Eunate Mayor. A logical analysis of responsibility attribution: emotions, individuals and collectives. *Journal of Logic and Computation*, 24(6):1313–1339, 2014.

[Lorini, 2013] Emiliano Lorini. Temporal STIT logic and its application to normative reasoning. *Journal of Applied Non Classical Logics*, 23(4):372–399, 2013.

[Schwarzentruber, 2012] François Schwarzentruber. Complexity results of STIT fragments. *Studia Logica*, 100(5):1001–1045, 2012.

[van der Hoek *et al.*, 2011] Wiebe van der Hoek, Nicolas Troquard, and Michael Wooldridge. Knowledge and control. In Liz Sonenberg, Peter Stone, Kagan Tumer, and Pinar Yolum, editors, *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 719–726. IFAAMAS, 2011.