ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-1093

STIT is dangerously undecidable

François Schwarzentruber[1] and **Caroline Semmling**[2]

Abstract. STIT is a potential logical framework to capture responsibility, counterfactual emotions and norms, which are main ingredients for specifying behaviors of virtual agents. We identify here a new fragment and its satisfiability problem is NP-complete and in Σ_3 when the number of agents is unbounded. We also identify a slightly more expressive fragment which is undecidable.

1 Motivation

In order to specify behaviors of artificial agents, it is essential to capture such concepts related to agency like knowledge, emotions, intentions and responsibilities. Hence STIT can be a promising candidate for the underlying necessary description of the agency. For instance, an agent a regrets that φ is true if a desires $\neg\varphi$ and he knows that φ is true and that other agents do not see to it that φ. Therefore we examine in this paper the STIT framework and focus on the satisfiability problem.

It should be recalled that the satisfiability problem of atemporal group STIT is undecidable and that the atemporal group STIT is not finitely axiomatizable, when the number of agents is greater than three [7]. The individual atemporal STIT,[3] however, is finitely axiomatizable [17] and, if the number of agents is greater than two, it is NEXPTIME-complete [1]. Recently, a variant called XSTIT [4, 5] has been proven to be decidable [14] (the semantics of XSTIT is not standard but given in terms of XSTITmodels).

In this paper, we provide new complexity results concerning the STIT framework. More precisely:

- We identify a new fragment whose satisfiability problem is NP-complete;
- We prove that if the number of agents is unbounded, then its satisfiability problem is in Σ_3;
- A slightly more expressive fragment is shown to be undecidable. This fragment is far from being the full language. The moral is the following: STIT is dangerously undecidable.

In section 2 we recall the semantics of STIT. In section 3 we discuss about fragments of STIT.

2 The STIT framework

In this section we define the language and the semantics of atemporal group STIT.

2.1 Syntax

The language of group STIT are built from a denumerable infinite set of propositional variables ATM and a finite set of n agents, $AGT = \{1, \ldots, n\}$. Well-formed formulas of the group STIT language are defined by:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [J]\varphi,$$

where $p \in ATM$ and $J \subseteq AGT$. The other Boolean connectives are defined as usual. The expression $[J]\varphi$ is read as "the group of agents J sees to it that φ". The expression $\langle J\rangle\varphi$ is introduced as abbreviation for $\neg[J]\neg\varphi$. It is read as "the group of agents J does not prevent φ" or "the particular choices of all agents in J do not rule out φ". It is conventional to write $[i]\varphi$ in place of $[\{i\}]\varphi$. The formula $[\emptyset]\varphi$ is read as "it is historically necessary that φ" and the formula $\langle\emptyset\rangle\varphi$ expresses that "it is historically possible that φ". The construction $\langle\emptyset\rangle[J]\varphi$ says it is possible that J sees to it that φ.

2.2 Semantics with product models

In spite of the philosophically well-founded background of BT+AC structures [2, 3, 8], an alternative semantics is proposed in [9] that is closer to standard model semantics. STIT is non-deterministic: there may be only several outcomes even if all agents have chosen an action. Deterministic STIT is variant of the original STIT where $[AGT]\varphi \leftrightarrow \varphi$ is a validity for all φ and where there is only one outcome if all agents have chosen an action. Another equivalent semantics for deterministic STIT is given given in terms of STIT-product models [12]. These models are defined in analogy to the *normal form* of games.

Definition 1 A STIT-product model is a tuple $\mathcal{M} = (W, V)$ defined as follows:

- $W = \Pi_i W_i$ where W_i are non-empty sets;
- $V : W \to 2^{ATM}$;

Accordingly, the truth condition of the group STIT operator is given by:

- $\mathcal{M}, w \models [J]\varphi$ iff for all $u \in W$ such that $u_j = w_j$ for all $j \in J$ we have $\mathcal{M}, u \models \varphi$.

With STIT-product models, we obtain the same set of satisfiable formulas that for deterministic STIT.

Actually, considering deterministic STIT is not in itself a restriction. Indeed, we obtain an algorithm for the satisfiability problem of non-deterministic STIT built from an algorithm for deterministic STIT because of the following embedding that is similar to the one presented in [6]:

[1] ENS Rennes, email: schwarze@ens-rennes.fr
[2] Institute of Philosophy II, Ruhr University Bochum, Germany, email: caroline.semmling@rub.de
[3] Some authors [4, 16] use the term 'multi-agent STIT' to designate the logic where operators are of the form $[i\ cstit :]$. We prefer here the use of the more explicit term 'individual STIT' as in [7] to distinguish individual from group STIT.

Theorem 1 *Let $AGT* = AGT \cup \{\infty\}$ where ∞ is a fresh new agent. Let φ a* STIT*-formula in which the set of agents is AGT. Then φ is satisfiable in a non-deterministic* STIT *model iff $tr_1(\varphi)$ is satisfiable in a deterministic* STIT*-model where tr is defined by:*

- $tr_1(p) = p$;
- $tr_1([J]\psi) = [J \cup \{\infty\}]tr_1(\psi)$.

Thus, in section 3, we only consider deterministic STIT.

3 Fragments

In this section, we investigate fragments given by the following grammar:

$$\begin{array}{ll}
\varphi ::= \chi \mid \psi \mid \varphi \wedge \varphi \mid \neg\varphi^{(1)} \mid \langle\emptyset\rangle\psi & \text{("can" formulas)} \\
\psi ::= [J]\chi \mid \psi \wedge \psi \mid \neg\psi^{(2)} & \text{("see-to-it" formulas)} \\
\chi ::= \bot \mid p \mid \chi \wedge \chi \mid \neg\chi & \text{(propositional formulas).}
\end{array}$$

where p ranges over atomic propositions and J ranges over subsets of agents. χ-formulas are propositional formulas. ψ-formulas are called "see-to-it" formulas and are Boolean formulas over formulas of the form $[J]\chi$ where J is a group of agents and χ is a Boolean formula. φ are Boolean formulas of χ-formulas and ψ-formulas but also constructions of the formula $\langle\emptyset\rangle\psi$. They are called "can" formulas.

The fragment $dfSTIT$ where negation constructions of type (1) are allowed but negation constructions of type (2) are not allowed has already been investigated in [11].

The fragment $dfSTIT2$ where negation constructions of type (2) are allowed but negation constructions of type (1) is the contribution of this article.

Finally we consider $boolSTIT$ where negation constructions of both type (1) and type (2) are both allowed.

We give complexity results concerning the satisfiability problem depending whether the number of agents is fixed (that is, we have a problem for each number of agents) or unbounded (the input can be any STIT-formula with as many agents we want). The following table sums up the results and the new results are written in bold.

Fragment	Fixed number of agents	Unbounded number of agents
$dfSTIT$	NP-complete	**NP-hard** **co-NP-hard** **in NEXPTIME**
$dfSTIT2$	**NP-complete**	**NP-hard** **co-NP-hard** **in the class Σ_3**
$boolSTIT$	**undecidable** **if more than 3 agents**	**undecidable**

where $\Sigma_3 = NP^{co-NP^{NP}}$. The class Σ_3 includes problems that can been solved by an alternating algorithm running in polynomial time where the alternation of quantifiers is $\exists\forall\exists$. For more information to Σ_3, one may refer to [13]. Proofs can be in [15].

4 Conclusion

The main concern of this paper deals with the satisfiability problem of the formulas over standard semantics of atemporal STIT framework. As it is wellknown, the atemporal group STIT logic and the product logic $\mathsf{S5}^n$ are related: one can polynomial reduce the satisfiability problem of $\mathsf{S5}^n$ to the satisfiability problem of atemporal group STIT logic [7] and vice versa.

So all results about $\mathsf{S5}^n$ can been transferred to atemporal STIT logic and vice versa, first of all the undecidability of the SAT problem. We have seen that there is a fine line between decidability and undecidability. Even though the modal depth is restricted to two, there are fragments of STIT which are undecidable.

Agi Kurucz shows that the product $\mathsf{S5}^3$ has not the finite model property [10] by exhibiting a formula with just four propositions. A model satisfying that formula has to have an infinite number of worlds. Consequently, atemporal group STIT also does not have the FMP. An open question, however, is the complexity of the satisfiability problem of atemporal group STIT when the number of proposition is restricted to three.

STIT can be a key feature of reasoners about counterfactual emotions, responsability, etc. As some fragments are NP-complete, a far good pratical approach would be to use SAT techniques. Another idea may be to develop efficient tableau method that can be integrated in already existing solvers that rely on this technique for modal and description logics.

ACKNOWLEDGEMENTS

We would like to thank the referees for their comments, which helped improve this paper considerably.

REFERENCES

[1] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard, 'Alternative axiomatics and complexity of deliberative STIT theories', *Journal of Philosophical Logic*, (2008).

[2] Nuel Belnap, Michael Perloff, and Ming Xu, *Facing the Future: Agents and Choices in Our Indeterminist World*, Oxford University Press, Oxford, 2001.

[3] Nuel Belnap and Micheal Perloff, 'Seeing to it that: a canonical form for agentives', *Theoria*, **54**, 175–199, (1988).

[4] J. Broersen, 'A complete stit logic for knowledge and action, and some of its applications', *Declarative Agent Languages and Technologies VI*, 47–59, (2009).

[5] J.M. Broersen, 'Deontic epistemic *stit* logic distinguishing modes of mens rea', *Journal of Applied Logic*, **9**(2), 127 – 152, (2011).

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

[7] A. Herzig and F. Schwarzentruber, 'Properties of logics of individual and group agency', in *Advances in Modal Logic*, pp. 133–149, (2008).

[8] 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).

[9] Barteld Kooi and Allard Tamminga, 'Moral conflicts between groups of agents', *Journal of Philosophical Logic*, **37**(1), 1–21, (2008).

[10] Á. Kurucz, 'S5 x s5 x s5 lacks the finite model property", *Advances in modal logic*, **3**, (2000).

[11] E. Lorini and F. Schwarzentruber, 'A logic for reasoning about counterfactual emotions', *Artificial Intelligence*, **175**(3), 814–847, (2011).

[12] Emiliano Lorini and François Schwarzentruber, 'A logic for reasoning about counterfactual emotions', in *IJCAI*, pp. 867–872, (2009).

[13] C.H. Papadimitriou, *Computational complexity*, John Wiley and Sons Ltd., 2003.

[14] Gillman Payette, 'Decidability of an xstit logic', *Studia Logica*, 1–31.

[15] F. Schwarzentruber and C. Semmling, 'Stit is dangerously undecidable', Technical report, INRIA, (2014).

[16] Heinrich Wansing, 'Tableaux for multi-agent deliberative-STIT logic', in *Advances in Modal Logic, Volume 6*, eds., Guido Governatori, Ian Hodkinson, and Yde Venema, 503–520, King's College Publications, (2006).

[17] Ming Xu, 'Axioms for deliberative STIT', *Journal of Philosophical Logic*, **27**, 505–552, (1998).