A STIT Logic Analysis of Social Influence

Emiliano Lorini
IRIT-CNRS
Toulouse University, France
lorini@irit.fr

Giovanni Sartor
CIRSFID, University of Bologna, Italy
European University Institute of Florence, Italy
giovanni.sartor@unibo.it

ABSTRACT

In this paper we propose a method for modeling social influence within the STIT approach to action. Our proposal consists in extending the STIT language with special operators that allows us to represent the consequences of an agent's choices over the rational choices of another agent.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal logic

Keywords

Logics of agency, STIT, social influence

1. INTRODUCTION

Both human and artificial societies are based on mutual influence. Agents are dependent on others for the realization of their goals, and only by influencing others, and obtaining their cooperation, they can adapt the physical and social world to their needs. Influence may take place through speech acts, as when one issues a request, an order or an advice, promises a reward to or threatens a sanction. It may also result from non-communicative behavior that, intentionally or unintentionally, obstacles or facilitates the performance of an action by another, as when one consumes a resource or blocks or limits access to a resource.

There have been a number of significant contributions to the logic of social influence (see in particular [18]) and to the cognitive aspects and computational aspects involved in it [5, 17]. However, no attempt has yet been done to capture influence in the framework of the STIT logic of action [2], though this logic presents some aspects which make it most promising for analyzing social influence. First of all, STIT naturally supports modeling the temporal aspect of influence, where the influencing action must precede the action being influenced. Secondly, STIT naturally supports addressing the strategic aspects of influencing relationships through extensive-form games.

In this paper we aim at showing that STIT can provide indeed a useful framework for modeling influence relationships. For this purpose, however STIT needs to be integrated with appropriate constructs, which make the influencee's agency consistent with the fact that the influencer determines the influencee's choices.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

The paper is organized as follows. In Section 2 we recall the general semantics of STIT, while in Section 3 we discuss the concept of influence from an informal perspective. Section 4 introduces a variant of STIT logic that will be used in Section 6 to formalize the concept of social influence informally discussed in Section 3. Our logic, called DR-STIT (STIT with Deterministic time and Rational choices), extends the basic STIT language with special operators that allow us to represent the consequences of an agent's *rational* choice. An axiomatization of DR-STIT is given in Section 5. Section 7 is about related work, while Section 8 discusses some perspectives of future research.

2. BACKGROUND ON STIT SEMANTICS

STIT logic (the logic of *seeing to it that*) by Belnap et al. [2] is one of the most prominent formal accounts of agency. It is the logic of sentences of the form "the agent i sees to it that φ is true". In [9] Horty extends Belnap et al.'s STIT framework with operators of group agency in order to express sentences of the form "the group of agents J sees to it that φ is true". Though also [2] approaches collective ('joint') agency, Horty's variant of group STIT is the most established today, and it provides the standard combination of agency operators for the individuals and agency operators for the groups. Different semantics for STIT have been proposed in the literature (see, e.g., [2, 4, 24, 15, 14, 22]). The original semantics of STIT by Belnap et al. [2] is defined in terms of $\mathbf{BT}+\mathbf{AC}$ structures: branching-time structures ($\mathbf{BT}$) augmented by agent choice functions ($\mathbf{AC}$). A $\mathbf{BT}$ structure is made of a set of moments and a tree-like ordering over them. An $\mathbf{AC}$ for an agent i is a function mapping each moment m into a partition of the set of histories passing through that moment, a history h being a maximal set of linearly ordered moments and the equivalence classes of the partition being the possible choices for agent i at moment m.

Following [14], here we adopt a Kripke-style semantics for STIT which has the advantage of being closer to the standard semantics of modal logic [3] than Belnap et al.'s original semantics. The main difference between a Kripke semantics for STIT and Belnap et al.'s $\mathbf{BT}+\mathbf{AC}$ semantics is that the former takes the concept of *world* as a primitive instead of the concept of *moment* and defines: (i) a *moment* as an equivalence class induced by a certain equivalence relation over the set of worlds, (ii) a *history* as a linearly ordered set of worlds induced by a certain partial order over the set of worlds, and (iii) an agent i's set of *choices* at a moment as a partition of that moment.

The Kripke semantics of STIT is illustrated in Figure 1, where each moment m_1, m_2 and m_3 consists of a set of worlds represented by points. For example, moment m_1 consists of the set of worlds $\{w_1, w_2, w_3, w_4\}$. Moreover, for every moment there exists a set of histories passing through it, where a history is defined

as a linearly ordered set of worlds. For example, the set of histories passing through moment m_1 is $\{h_1, h_2, h_3, h_4\}$. Finally, for every moment, there exists a partition which characterizes the set of available choices of agent 1 in this moment. For example, at moment m_1, agent 1 has two choices, namely $\{w_1, w_2\}$ and $\{w_3, w_4\}$. Note that an agent's set of choices at a certain moment can also be seen as a partition of the set of histories passing through this moment. For example, we can identify the choices available to agent 1's at m_1 with the two sets of histories $\{h_1, h_2\}$ and $\{h_3, h_4\}$.

Clearly, for every moment m in a Kripke semantics for STIT, one can identify the set of histories passing through it by considering all histories that contain at least one world in the moment m. Moreover, an agent i's set of choices available at m can also be seen as as a partition of the set of histories passing through m. At first glance, an important difference between Belnap et al.'s semantics and Kripke semantics for STIT seems to be that in the former the truth of a formula is relative to a moment-history pair m/h, also called *index*, whereas in the latter it is relative to a world w. However, this difference is only apparent, because in the Kripke semantics for STIT there is a one-to-one correspondence between worlds and indexes, in the sense that: (i) for every index m/h there exists a unique world w at the intersection between m and h, (ii) and for every world w there exists a unique index m/h such that the intersection between m and h includes w. (This point will become clearer in Section 4.2 in which a Kripke semantics for our variant of STIT will be specified.)

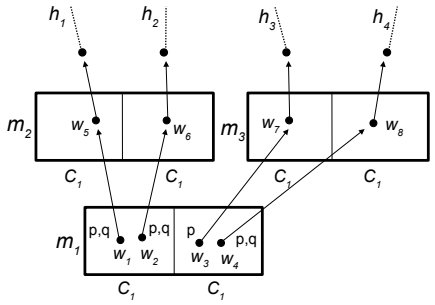

Figure 1: Illustration of Kripke semantics of STIT

The STIT semantics provides for different concepts of agency, all characterized by the fact that an agent acts only if she sees to it that a certain state of affairs is the case. In this paper we consider two different notions of agency, namely the so-called Chellas STIT, named after its proponent [6], and the deliberative STIT [10].[1]

An agent i Chellas-sees-to-it that φ, denoted by formula $[i\ \mathbf{stit}]\varphi$, at a certain world w if and only if, for every world v, if w and v belong to the same choice of agent i then φ is true at world v. For example, in Figure 1, agent 1 Chellas-sees-to-it that p at world w_1 because p is true both at world w_1 and at at world w_2.

Deliberative STIT satisfies the same positive condition as Chellas STIT *plus* a negative condition: an agent i deliberatively-sees-to-it that φ, denoted by formula $[i\ \mathbf{dstit}]\varphi$, at a certain world w if and only if: (i) agent i Chellas-sees-to-it that φ at w, and (ii) there exists a world v such that w and v belong to the same moment and φ is false at v. For example, in Figure 1, agent 1 deliberatively sees to it that q at world w_1 because q is true both at world w_1 and at world w_2, while being false at world w_3. In other terms, while the truth of $[i\ \mathbf{stit}]\varphi$ only requires that i's choice ensures that φ, the truth of $[i\ \mathbf{dstit}]\varphi$ also requires that i had the opportunity of making an alternative choice that would not guarantee that φ would be the case. Deliberative STIT, we would argue, captures a fundamental aspect of the concept of action, namely, the idea that for a state of affairs to be the consequence of an action (or for an action to be the cause of a state of affairs), it is not sufficient that the action is a sufficient condition for that state of affairs to hold, it is also required that, without the action, the state of affairs possibly would not hold (a similar idea is also included in the logic of "bringing it about" by Pörn, see in particular [19]). In this sense, while $[i\ \mathbf{stit}]\varphi$ at w is consistent with (and is indeed entailed by) the necessity of φ at w, $[i\ \mathbf{dstit}]\varphi$ at w is incompatible with the necessity of φ at w, since it requires that at w also $\neg\varphi$ was an open possibility. Consequently, the deliberative STIT is more appropriate than the Chellas STIT to describe the consequences of an agent's action, as *incompatibility with necessity* is a requirement for any reasonable concept of action.[2]

[1]We shall not consider achievement STIT of [1].

3. THE CONCEPT OF INFLUENCE

Our analysis of social influence starts from G. E. Moore's famous view that free will and voluntariness are compatible with determinism, that is to say, the fact that a voluntary action could be determined by some external causes. According to Moore's famous analysis [16], one is free in performing an action if one "could have done otherwise", but the latter expression has to be understood in a particular way, namely, as the requirement that "should one have done otherwise if one had chosen to do so". Thus this kind of freedom is consistent with the view that the choice of an agent is determined by the agent's nature, namely, by the agent's preferences and rationality, and more generally by the way in which the agent's decisional process works. It is also consistent with the idea that the external conditions in which an agent finds herself or the other agents with whom the agent interacts may provide an input to the agent's decision-making process in such a way that a determinate choice should follow. As Leibniz [12, 383] observed, voluntary action require the will, they will happen "because one will do, and because one will will to do, that which leads to them". However, the formation of such a will may be influenced by external causes. In particular, "precepts, armed with power to punish and to recompense, are very often of use and are included in the order of causes which make an action exist".

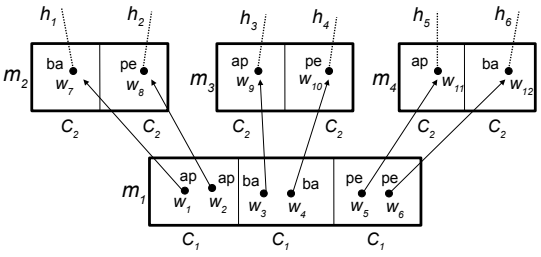

Figure 2: Branching fruits.

We argue that genuine influence consists in *determining* the voluntary action of an agent by modifying her choice set, so that a different choice becomes preferable to the influencee on comparison to what would be her preferred option without this modification. This may happen by expanding or restricting the set of the available choices, or by changing the payoffs associated to such

[2]The classical argument against the use of Chellas STIT for modeling action is that, according to Chellas STIT, an agent brings about all tautologies and that it is counterintuitive to say that a tautology is a consequence of an agent's action.

choices (as when rewards or punishments are established). To illustrate this concept of influence, let us consider the example in Figure 2. The example represents a situation where there are three fruits on a table, an apple, a banana and a pear. The actions at issue consist in bringing about that the apple is eaten (ap), the banana is eaten (ba) or the pear is eaten (pe). Let us assume that agent 2 has certain preferences: *she prefers eating apples to bananas to pears* ($ap > ba > pe$). Let us also assume that 2 is rational, in the sense that she acts in such a way as to achieve the outcome she prefers. By choosing to eat the apple at w_1, 1 generates a situation where, given her preferences, 2 will necessarily eat the banana, rather than the pear. Indeed, although at moment m_2, 2 has two choices available, namely the choice of eating the banana and the choice of eating the pear, only the former is rational, in the sense of being compatible with 2's preferences. In this sense, by deciding to eat the apple at w_1, 1 influences 2 to decide to eat the banana at w_7. This example leads us to the following informal definition of social influence:

> An agent i influences another agent j to perform a certain (voluntary) action if and only if, i sees to it that that every rational choice of j will lead j to perform the action.

In the next section we present a logic which enables us to formalize the previous concept of social influence. Specifically, it enables us to represent both aspects of social influence relations: the influencee's freedom to select the action she prefers in her choice set, and the influencer's ability of determining the influencee's choice by modifying the influencee's choice set.

4. DR-STIT LOGIC

Our logic is a variant of STIT with discrete time and rational choices interpreted in Kripke semantics. We call DR-STIT (STIT with Deterministic time and Rational choices) this logic. On the syntactic level, DR-STIT is nothing but the extension of atemporal individual STIT by: (i) the temporal operators 'next' (tomorrow) and 'previous' (yesterday) of linear temporal logic, (ii) the operator of group agency for the *grand coalition* (the coalition of all agents), and (iii) special operators of agency describing the consequences of an agent's rational choice.

In DR-STIT the so-called Chellas STIT operators are taken as primitive operators of agency. As pointed out by [10], deliberative STIT operators and Chellas STIT operators are interdefinable and just differ in the choice of primitive operators.

The following two sections present the syntax and a Kripke-style semantics for DR-STIT (Subsections 4.1 and 4.2).

4.1 Syntax

Assume a countable (possibly infinite) set of atomic propositions denoting facts $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$.

The language $\mathcal{L}_{\text{DR-STIT}}(Atm, Agt)$ of the logic DR-STIT is the set of formulae defined by the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid \mathbf{X}\varphi \mid \mathbf{Y}\varphi \mid [i\ \mathbf{stit}]\varphi \mid [Agt\ \mathbf{stit}]\varphi \mid [i\ \mathbf{rstit}]\varphi$$

where p ranges over Atm and i ranges over Agt. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from $\neg$ and $\wedge$ in the standard way.

Operators of the form $[i\ \mathbf{stit}]$ are Chellas STIT operators that have been informally discussed in Section 2. Formula $[i\ \mathbf{stit}]\varphi$ captures the fact that φ is guaranteed by a present action of agent i, and has to be read 'agent i sees to it that φ regardless of what the other agents do'. We shorten the reading of $[i\ \mathbf{stit}]\varphi$ to 'agent i sees to it that φ'. $[Agt\ \mathbf{stit}]$ is a group STIT operator which captures the fact that φ is guaranteed by a present choice of all agents, and has to be read 'all agents see to it that φ by acting together'. The modal operator $[Agt\ \mathbf{stit}]$ will be fundamental in Section 5 in order to axiomatize a basic property relating action and time studied in STIT: the so-called property of *no choice between undivided histories* [2, Chap. 7]. The dual operators of $[i\ \mathbf{stit}]$ and $[Agt\ \mathbf{stit}]$ are defined in the usual way:

$$\langle i\ \mathbf{stit}\rangle\varphi \ \stackrel{\text{def}}{=}\ \neg[i\ \mathbf{stit}]\neg\varphi,$$

$$\langle Agt\ \mathbf{stit}\rangle\varphi \ \stackrel{\text{def}}{=}\ \neg[Agt\ \mathbf{stit}]\neg\varphi.$$

Operators of the form $[i\ \mathbf{rstit}]$ describe the effects of a rational (or preferred) choice of agent i. Following one of the arguments of Section 3, namely the idea that a choice is rational if it is consistent with the agent's preference over her alternative choices, we here conceive the terms 'rational choice' and 'preferred choice' as synonyms. Specifically, formula $[i\ \mathbf{rstit}]\varphi$ has to be read either "if agent i's current action is the result of a rational choice of i, then i sees to it that φ" or, "if agent i's current choice is a preferred choice of i, then i sees to it that φ".[3] The dual of the operator $[i\ \mathbf{rstit}]$ is defined as follows: $\langle i\ \mathbf{rstit}\rangle\varphi \stackrel{\text{def}}{=} \neg[i\ \mathbf{rstit}]\neg\varphi$. Note that $\langle i\ \mathbf{rstit}\rangle\top$ has to be read either "agent i's current choice is rational" or, "agent i's current choice is a preferred choice of i". The formula $[i\ \mathbf{stit}]\varphi \wedge \langle i\ \mathbf{rstit}\rangle\top$, which is logically equivalent to $[i\ \mathbf{rstit}]\varphi \wedge \langle i\ \mathbf{rstit}\rangle\top$, has to be read either "agent i sees to it that φ as a result of her rational choice" or, "agent i sees to it that φ as a result of her preferred choice".

$\Box\varphi$ stands for 'φ is true regardless of what every agent does' or 'φ is true no matter what the agents do' or simply 'φ is necessarily true'. We define the dual of $\Box$ as follows: $\Diamond\varphi \stackrel{\text{def}}{=} \neg\Box\neg\varphi$. Note that the operators $[i\ \mathbf{stit}]$ and $\Box$ can be combined in order define the deliberative STIT operator $[i\ \mathbf{dstit}]$ we have discussed in Section 2:

$$[i\ \mathbf{dstit}]\varphi \ \stackrel{\text{def}}{=}\ [i\ \mathbf{stit}]\varphi \wedge \neg\Box\varphi.$$

Moreover, the operators $[i\ \mathbf{rstit}]$ and $\Box$ can be combined in order define a special kind of deliberative STIT operator for rational choices:

$$[i\ \mathbf{rdstit}]\varphi \ \stackrel{\text{def}}{=}\ \langle i\ \mathbf{rstit}\rangle\top \rightarrow [i\ \mathbf{dstit}]\varphi.$$

$[i\ \mathbf{rdstit}]\varphi$ has to be read either "if agent i's current action is the result of a rational choice of i, then i *deliberatively* sees to it that φ" or, "if agent i's current choice is a preferred choice of i, then i *deliberatively* sees to it that φ".

Finally, $\mathbf{X}$ and $\mathbf{Y}$ are the standard operators 'next' (tomorrow) and 'previous' (yesterday) of linear temporal logic. $\mathbf{X}\varphi$ and $\mathbf{Y}\varphi$ has to be read respectively 'φ is going to be true in the next world' and 'φ was true in the previous world'.

4.2 Kripke semantics for DR-STIT

The basic notion in the semantics is the notion of Kripke STIT model with discrete time and rational choices. For notational convenience, in what follows we are going to use the following abbreviations. Given a set of elements W, an arbitrary binary relation $\mathcal{R}$ on W and an element w in W, let $\mathcal{R}(w) = \{v \in W | w\mathcal{R}v\}$. Moreover, given two binary relations $\mathcal{R}_1$ and $\mathcal{R}_2$ on W let $\mathcal{R}_1 \circ \mathcal{R}_2$ be

[3]Our notion of "doing something as the result of a rational (or preferred) choice" is synonym of List & Rabinowicz's notion of "doing an action with endorsement" [13].

the standard operation of composition between binary relations. If we abstract away from the notion of rational choice, Kripke STIT models with discrete time and rational choices are nothing but a subclass of the general class of temporal Kripke STIT models studied by [14]. The latter can be seen as extensions of Zanardo's Ockhamist models [26] by a choice component, i.e., by accessibility relations for individual choices and an accessibility relation for the collective choice of the grand coalition Agt.

DEFINITION 1. *A Kripke STIT model with discrete time and rational choices is a tuple* $M = (W, \equiv, \rightarrow\!\triangleright, \triangleleft\!\leftarrow, \{\mathcal{C}_i\}_{i\in Agt}, \mathcal{C}_{Agt}, \{\mathbf{RC}_i\}_{i\in Agt}, \mathcal{V})$ *where:*

- W *is a nonempty set of possible worlds;*
- $\equiv$ *is an equivalence relation on* W*;*
- $\rightarrow\!\triangleright$ *is a serial and deterministic relation on* W*;*
- $\triangleleft\!\leftarrow$ *is the inverse relation of* $\rightarrow\!\triangleright$ *(i.e.,* $\triangleleft\!\leftarrow = \{(w,v)|v \rightarrow\!\triangleright w\}$*) and is supposed to be deterministic;*
- $\mathcal{C}_{Agt}$ *and every* $\mathcal{C}_i$ *are equivalence relations on* W*;*
- *every* $\mathbf{RC}_i$ *is a subset of the partition of* W *induced by the equivalence relation* $\mathcal{C}_i$*;*
- $\mathcal{V} : Atm \longrightarrow 2^W$ *is a valuation function for atomic propositions;*

and that satisfies the following six constraints:

(C1) *for all* $w \in W$*: if* $w\mathcal{F}v$ *then* $w \not\equiv v$*, with* $\mathcal{F}$ *denoting the transitive closure of the binary relation* $\rightarrow\!\triangleright$*;*

(C2) *for all* $i \in Agt$*:* $\mathcal{C}_i \subseteq \equiv$*;*

(C3) *for all* $u_1, \ldots, u_n \in W$*: if* $u_i \equiv u_j$ *for all* $i, j \in \{1, \ldots, n\}$ *then* $\bigcap_{1\leq i\leq n} \mathcal{C}_i(u_i) \neq \emptyset$*;*

(C4) *for all* $w \in W$ *and for all* $i \in Agt$*: there exists* $v \in W$ *such that* $w \equiv v$ *and* $\mathcal{C}_i(v) \in \mathbf{RC}_i$*;*

(C5) *for all* $w \in W$*:* $\mathcal{C}_{Agt}(w) = \bigcap_{i\in Agt} \mathcal{C}_i(w)$*;*

(C6) $\rightarrow\!\triangleright \circ \equiv \subseteq\ \mathcal{C}_{Agt} \circ \rightarrow\!\triangleright$.

Let us explain in detail each component of the preceding definition.

$\equiv(w)$ is the set of worlds that are alternative to the world w. Following the Ockhamist's view of time [20, 26], we call the equivalence classes induced by the equivalence relation $\equiv$ *moments*. The set of all moments in the model M is denoted by Mom and the elements in Mom are denoted by $m, m', \ldots$

$\rightarrow\!\triangleright(w)$ is the set of direct temporal successors of world w, that is to say, $w \rightarrow\!\triangleright v$ means that v is in the future of w and there is no third world that is in the future of w and in the past of v. The fact that $\rightarrow\!\triangleright$ is serial and deterministic means that every world has *exactly one* direct temporal successor. $\triangleleft\!\leftarrow(w)$ defines the set of direct temporal predecessors of world w, that is to say, $v \triangleleft\!\leftarrow w$ means that v is in the past of w and there is no third world that is in the past of w and in the future of v. The fact that $\triangleleft\!\leftarrow$ is deterministic means that every world has *at most one* direct temporal predecessor. We do not assume $\triangleleft\!\leftarrow$ to be serial because past is not necessarily endless.

The Constraint C1 in Definition 1 ensures that if two worlds belong to the same moment then one of them cannot be in the future of the other. (Note that $\mathcal{F}(w)$ is the set of worlds that are in the future of w.) Since the relation $\equiv$ is reflexive, the Constraint C1 implies that the relations $\rightarrow\!\triangleright$, $\triangleleft\!\leftarrow$ and $\mathcal{F}$ are all irreflexive.

Let $\mathcal{T}(w) = \mathcal{P}(w) \cup \{w\} \cup \mathcal{F}(w)$ be the set of worlds that are temporally related with world w, where $\mathcal{P} = \{(w,v)|v\mathcal{F}w\}$ is the inverse of the relation $\mathcal{F}$ and $\mathcal{P}(w)$ is the set of worlds that are in the past of w. The fact that the relation $\mathcal{F}$ is irreflexive and transitive ensures that $\mathcal{F}$ is a strict linear (or total) order on the set $\mathcal{T}(w)$. For every world w in W, we call the linearly ordered set $(\mathcal{T}(w), \mathcal{F})$ the history going through w. Note that, because of the seriality of the relation $\rightarrow\!\triangleright$, every history is infinite. For notational convenience, let $Hist$ denote the set of all histories in the model M and let the elements of $Hist$ be denoted by $h, h', \ldots$ Moreover, for every moment $m \in Mom$, let

$$Hist_m = \{h \in Hist | \exists w \in W \text{ such that } w \in m \cap h\}$$

be the set of all histories passing through the moment m and let

$$Ind = \{m/h | m \in Mom \text{ and } h \in Hist_m\}$$

be the set of all indexes in the model M.

Clearly, our Kripke semantics for DR-STIT allows us to interchangeably use the term 'world' and 'history going through a certain world' without lost of generality. Indeed, for every world w there exists a unique history going through it. This point is highlighted by the following proposition.

PROPOSITION 1. *Let* $w \in W$*. Then, there exists a unique* $h \in Hist$ *such that* $w \in h$*.*

As every world in a model is identified with a unique history going through it, the equivalence relation $\equiv$ can also be understood as an equivalence relation between historic alternatives: $w \equiv v$ means that the history going through v is alternative to the history going through w, or the history going through w and the history going through v pass through the same moment.

Furthermore, in the semantics for DR-STIT there is a one-to-one correspondence between worlds and indexes, as for every index m/h in Ind there exists a unique world w at the intersection between m and h, and for every world w there exists a unique index m/h such that the intersection between m and h includes w. This fact is highlighted by the following two propositions.

PROPOSITION 2. *Let* $m \in Mom$ *and let* $h \in Hist_m$*. Then,* $m \cap h$ *is a singleton.*

PROPOSITION 3. *Let* $w \in W$*. Then, there exists a unique* $m/h \in Ind$ *such that* $w \in m \cap h$*.*

For every world w, the set $\mathcal{C}_i(w)$ identifies agent i's *actual* choice at w, that is to say, the set of worlds that can be obtained by agent i's actual choice at w. Because of the one-to-one correspondence between worlds and histories, one can also identify i's *actual* choice at w with the set $\{h \in Hist | \exists v \in \mathcal{C}_i(w) \text{ such that } v \in h\}$. In other words, in DR-STIT an agent chooses among different sets of histories.

Constraint C2 in Definition 1 just means that an agent can only choose among possible alternatives. This constraint ensures that, for every world w, the equivalence relation $\mathcal{C}_i$ induces a partition of the set $\equiv(w)$. An element of this partition is a choice that is *possible (or available)* for agent i at w.

Constraint C3 expresses the so-called assumption of *independence of agents* or *independence of choices*: if $\mathcal{C}_1(u_1)$ is a possible choice for agent 1 at w, $\mathcal{C}_2(u_2)$ is a possible choice for agent 2 at w,..., $\mathcal{C}_n(u_n)$ is a possible choice for agent n at w, then their intersection is non-empty. More intuitively, this means that agents can never be deprived of choices due to the choices made by other agents.

Let $\mathbf{C}_i$ denote the partition of the set of worlds W induced by the equivalence relation $\mathcal{C}_i$. This partition characterizes agent i's set of choices in the model M. The set $\mathbf{RC}_i \subseteq \mathbf{C}_i$ characterizes

agent i's set of *rational* choices or also, the set of *preferred* choices of agent i (given the assumption that a choice is rational if it is consistent with the agent's preference over her alternative choices). The Constraint C4 just means that, at each moment, an agent has at least one rational choice available.

For every world w, the set $\mathcal{C}_{Agt}(w)$ identifies the *actual* choice of group Agt at w, that is to say, the set of worlds that can be obtained by the collective choice of all agents at w. Constraint C5 just says that the collective choice of the grand coalition Agt is equal to the intersection of the choices of all individuals. This corresponds to the notion of joint action proposed by Horty in [9], where the joint action of a group is described in terms of the result that the agents in the group bring about by acting together.

The Constraint C6 expresses a basic relation between action and time: if v is in the future of w and u and v are in the same moment, then there exists an alternative z in the collective choice of all agents at w such that u is in the future of z. This constraint corresponds to the property of *no choice between undivided histories* given in STIT logic [2, Chap. 7]. It captures the idea that if two histories come together in some future moment then, in the present, each agent does not have a choice between these two histories. This implies that if an agent can choose between two histories at a later stage, then she does not have a choice between them in the present.

A formula φ of the logic DR-STIT is evaluated with respect to a given Kripke STIT model with discrete time and rational choices $M = (W, \equiv, \rightarrow, \leftarrow, \{\mathcal{C}_i\}_{i\in Agt}, \mathcal{C}_{Agt}, \{\mathbf{RC}_i\}_{i\in Agt}, \mathcal{V})$ and a world w in M. We write $M, w \models \varphi$ to mean that φ is true at world w in M. The truth conditions of DR-STIT formulae are then defined as follows:

$$
\begin{array}{rcl}
M, w \models p & \iff & w \in \mathcal{V}(p) \\
M, w \models \neg\varphi & \iff & M, w \not\models \varphi \\
M, w \models \varphi \wedge \psi & \iff & M, w \models \varphi \text{ AND } M, w \models \psi \\
M, w \models \Box\varphi & \iff & \forall v \in \equiv(w) : M, v \models \varphi \\
M, w \models \mathbf{X}\varphi & \iff & \forall v \in \rightarrow(w) : M, v \models \varphi \\
M, w \models \mathbf{Y}\varphi & \iff & \forall v \in \leftarrow(w) : M, v \models \varphi \\
M, w \models [i\ \mathbf{stit}]\varphi & \iff & \forall v \in \mathcal{C}_i(w) : M, v \models \varphi \\
M, w \models [Agt\ \mathbf{stit}]\varphi & \iff & \forall v \in \mathcal{C}_{Agt}(w) : M, v \models \varphi \\
M, w \models [i\ \mathbf{rstit}]\varphi & \iff & \text{IF } \mathcal{C}_i(w) \in \mathbf{RC}_i \text{ THEN} \\
 & & \forall v \in \mathcal{C}_i(w) : M, v \models \varphi
\end{array}
$$

For any formula φ of the language $\mathcal{L}_{\text{DR-STIT}}(Atm, Agt)$, we write $\models_{\text{DR-STIT}} \varphi$ if φ is DR-STIT *valid*, i.e., for all Kripke STIT models with discrete time and rational choices M and for all worlds w in M, we have $M, w \models \varphi$. We say that φ is DR-STIT *satisfiable* if $\neg\varphi$ is not DR-STIT valid.

5. AXIOMATIZATION

Figure 3 contains a complete axiomatization of the logic DR-STIT with respect to the class of Kripke STIT models with discrete time and rational choices. This includes all tautologies of classical propositional calculus **(PC)** as well as modus ponens **(MP)**. Moreover, we have all principles of the normal modal logic S5 for every operator $[i\ \mathbf{stit}]$, for the operator $[Agt\ \mathbf{stit}]$ and for the operator $\Box$, all principles of the normal modal logic KD for the temporal operator $\mathbf{X}$ and all principles of the normal modal logic K for the temporal operator $\mathbf{Y}$. That is, we have Axiom K for each operator: $(\blacksquare\varphi \wedge \blacksquare(\varphi \rightarrow \psi)) \rightarrow \blacksquare\psi$ with $\blacksquare \in \{\Box, \mathbf{X}, \mathbf{Y}, [Agt\ \mathbf{stit}]\} \cup \{[i\ \mathbf{stit}] | i \in Agt\}$. We have Axiom D for the temporal operator $\mathbf{X}$: $\neg(\mathbf{X}\varphi \wedge \mathbf{X}\neg\varphi)$. We have Axiom 4 for $\Box$, $[Agt\ \mathbf{stit}]$ and for every $[i\ \mathbf{stit}]$: $\blacksquare\varphi \rightarrow \blacksquare\blacksquare\varphi$ with $\blacksquare \in \{\Box, [Agt\ \mathbf{stit}]\} \cup \{[i\ \mathbf{stit}] | i \in Agt\}$. Furthermore, we have Axiom T for $\Box$, $[Agt\ \mathbf{stit}]$ and for every $[i\ \mathbf{stit}]$: $\blacksquare\varphi \rightarrow \varphi$ with $\blacksquare \in \{\Box, [Agt\ \mathbf{stit}]\} \cup \{[i\ \mathbf{stit}] | i \in Agt\}$. We have Axiom B for $\Box$, $[Agt\ \mathbf{stit}]$ and for every $[i\ \mathbf{stit}]$: $\varphi \rightarrow \blacksquare\neg\blacksquare\neg\varphi$ with $\blacksquare \in \{\Box, [Agt\ \mathbf{stit}]\} \cup \{[i\ \mathbf{stit}] | i \in Agt\}$. Finally we have the rule of necessitation for each modal operator: $\frac{\varphi}{\blacksquare\varphi}$ with $\blacksquare \in \{\Box, [Agt\ \mathbf{stit}], \mathbf{X}, \mathbf{Y}\} \cup \{[i\ \mathbf{stit}] | i \in Agt\}$.

We have principles for the temporal operators and for the relationship between time and action. **($\mathbf{Alt_X}$)** and **($\mathbf{Alt_Y}$)** are the basic axioms for the determinism of 'tomorrow' and 'yesterday'. **($\mathbf{Conv_{X,Y}}$)** and **($\mathbf{Conv_{Y,X}}$)** are the basic interaction axioms between 'tomorrow' and 'yesterday' according to which "if φ is true in the present, then tomorrow is going to be true that yesterday φ has been true" and "if φ is true in the present, then yesterday has been true that tomorrow φ is going to be true".

($\mathbf{Rel}_{\Box,[i\ \mathbf{stit}]}$) and **(AIA)** are the two central principles in Xu's axiomatization of the Chellas STIT operators $[i\ \mathbf{stit}]$ [25]. According to Axiom **($\mathbf{Rel}_{\Box,[i\ \mathbf{stit}]}$)**, if φ is true regardless of what every agent does, then every agent sees to it that φ. In other words, an agent brings about those facts that are inevitable.According to Axiom **($\mathbf{Rel}_{[i\ \mathbf{stit}],[Agt\ \mathbf{stit}]}$)**, all agents bring about together what each of them brings about individually.

Axiom **(NCUH)** establishes a fundamental relationship between action and time and corresponds to the semantic constraint of 'no choice between undivided histories' (Constraint C6 in Definition 1): if in the next world φ is going to be possible then the actual collective choice of all agents will possibly result in a world in which φ is true.

Axiom **($\mathbf{Rel}_{[i\ \mathbf{rstit}],[i\ \mathbf{stit}]}$)** is the basic interaction principle between the rational choice operator $[i\ \mathbf{rstit}]$ and the Chellas STIT operator $[i\ \mathbf{stit}]$. Axiom **(RatCh)** means that the rationality of an agent just depends on her actual choice: if agent i is rational then i sees to it that she is rational. Finally, according to Axiom **(OneRat)**, an agent always has a rational choice in her repertoire.

THEOREM 1. *The set of DR-STIT validities is completely axiomatized by the principles given in Figure 3.*

PROOF. (Sketch) Space restrictions prevent us from giving a detailed proof of Theorem 1. Thus, we only give the general idea of the proof.

Proving that the principles given in Figure 3 are sound with respect to the class of Kripke STIT models with discrete time and rational choices (KDRs for short) is just a routine task. The proof of completeness requires more work and is divided in four steps.

From KDRs to standard KDRs. The first step consists in providing a DR-STIT semantics in terms of *standard* KDRs (SKDRs for short), that is, tuples of the form $(W, \equiv, \rightarrow, \leftarrow, \{\mathcal{C}_i\}_{i\in Agt}, \mathcal{C}_{Agt}, \{\mathcal{RC}_i\}_{i\in Agt}, \mathcal{V})$ where $W, \equiv, \leftarrow, \rightarrow, \{\mathcal{C}_i\}_{i\in Agt}, \mathcal{C}_{Agt}$ and $\mathcal{V}$ are exactly as in Definition 1 and $\{\mathcal{RC}_i\}_{i\in Agt}$ is a family of binary relations on W that satisfy the following three constraints for all $i \in Agt$:

(C7) for all $w \in W$: if $\mathcal{RC}_i(w) \neq \emptyset$ then $\mathcal{RC}_i(w) = \mathcal{C}_i(w)$;

(C8) for all $w, v \in W$: if $v \in \mathcal{C}_i(w)$ and $\mathcal{RC}_i(w) \neq \emptyset$ then $\mathcal{RC}_i(v) \neq \emptyset$;

(C9) for all $w \in W$: there exists $v \in W$ such that $w \equiv v$ and $\mathcal{RC}_i(v) \neq \emptyset$.

It is easy to prove that:

LEMMA 1. *For every formula φ in $\mathcal{L}_{\text{DR-STIT}}(Atm, Agt)$, φ is satisfiable relative to the class of KDRs iff it is satisfiable relative to the class of SKDRs.*

PC	Tautologies of classical propositional calculus
S5($\Box$)	All S5-principles for the operator $\Box$
KD(X)	All KD-principles for the operator $\mathbf{X}$
K(Y)	All K-principles for the operator $\mathbf{Y}$
S5($[i\ \mathbf{stit}]$)	All S5-principles for the operators $[i\ \mathbf{stit}]$
S5($[Agt\ \mathbf{stit}]$)	All S5-principles for the operator $[Agt\ \mathbf{stit}]$
(Alt$_\mathbf{X}$)	$\neg\mathbf{X}\varphi \rightarrow \mathbf{X}\neg\varphi$
(Alt$_\mathbf{Y}$)	$\neg\mathbf{Y}\varphi \rightarrow \mathbf{Y}\neg\varphi$
(Conv$_{\mathbf{X},\mathbf{Y}}$)	$\varphi \rightarrow \mathbf{XY}\varphi$
(Conv$_{\mathbf{Y},\mathbf{X}}$)	$\varphi \rightarrow \mathbf{YX}\varphi$
(Rel$_{\Box,[i\ \mathbf{stit}]}$)	$\Box\varphi \rightarrow [i\ \mathbf{stit}]\varphi$
(AIA)	$(\Diamond[1\ \mathbf{stit}]\varphi_1 \wedge \ldots \wedge \Diamond[n\ \mathbf{stit}]\varphi_n) \rightarrow$ $\Diamond([1\ \mathbf{stit}]\varphi_1 \wedge \ldots \wedge [n\ \mathbf{stit}]\varphi_n)$
(Rel$_{[i\ \mathbf{rstit}],[i\ \mathbf{stit}]}$)	$\langle i\ \mathbf{rstit}\rangle\top \rightarrow ([i\ \mathbf{rstit}]\varphi \leftrightarrow [i\ \mathbf{stit}]\varphi)$
(RatCh)	$\langle i\ \mathbf{rstit}\rangle\top \rightarrow [i\ \mathbf{stit}]\langle i\ \mathbf{rstit}\rangle\top$
(OneRat)	$\Diamond\langle i\ \mathbf{rstit}\rangle\top$
(Rel$_{[i\ \mathbf{stit}],[Agt\ \mathbf{stit}]}$)	$([1\ \mathbf{stit}]\varphi_1 \wedge \ldots \wedge [n\ \mathbf{stit}]\varphi_n) \rightarrow$ $[Agt\ \mathbf{stit}](\varphi_1 \wedge \ldots \wedge \varphi_n)$
(NCUH)	$\mathbf{X}\Diamond\varphi \rightarrow \langle Agt\ \mathbf{stit}\rangle\mathbf{X}\varphi$
(MP)	$\dfrac{\varphi,\ \varphi \rightarrow \psi}{\psi}$

Figure 3: Axiomatization of DR-STIT

From SKDRs to SKDRs with possible cycles. The second step consists in defining a DR-STIT semantics in terms of standard Kripke STIT models with discrete time, rational choices *and possible cycles* (SKDRCs). The latter are like SKDRs except that they do not necessarily satisfy the Constraint C1 in Definition 1 about the special kind of irreflexivity of the temporal relation $\mathcal{F}$ between moments. We prove that:

LEMMA 2. *For every formula φ in $\mathcal{L}_{DR\text{-}STIT}(Atm, Agt)$, φ is satisfiable relative to the class of SKDRs iff it is satisfiable relative to the class of SKDRCs.*

The lemma is provable by adapting the method used in [8, Th. 2] which allows us to show that, for every SKDRC M, we can build a SKDR M' and define a bounded morphism from M' to M [3, Def. 2.12]. In other words, we show that the Constraint C1 is not modally definable in the logic DR-STIT.

From SKDRCs to superadditive SKDRCs. The third step consists in introducing a DR-STIT semantics in terms of *superadditive* SKDRCs. The only difference between SKDRCs and *superadditive* SKDRCs is that in the latter the Constraint C5 in Definition 1 is replaced by the following weaker Constraint C5*:

(C5*) for all $w \in W$: $\mathcal{C}_{Agt}(w) \subseteq \bigcap_{i\in Agt} \mathcal{C}_i(w)$.

We prove that:

LEMMA 3. *For every formula φ in $\mathcal{L}_{DR\text{-}STIT}(Atm, Agt)$, φ is satisfiable relative to the class of SKDRCs iff it is satisfiable relative to the class of* superadditive *SKDRCs.*

The lemma is provable by adapting the method used in [14, Lemma 1] which allows us to show that, for every SKDRC M, we can build a superadditive SKDRC M' and define a bounded morphism from M' to M. In other words, we show that the direction $\mathcal{C}_{Agt}(w) \supseteq \bigcap_{i\in Agt} \mathcal{C}_i(w)$ of the Constraint C5 is not modally definable in the logic DR-STIT.

Completeness wrt superadditive SKDRCs. The fourth step consists in proving that the set of DR-STIT formulae that are valid in the class of superadditive SKDRCs is completely axiomatized by the principles given in Figure 3. Indeed, it is a routine task to check that all principles in Figure 3 correspond one-to-one to their semantic counterparts on the class of superadditive SKDRCs. In particular, **S5**($\Box$), **S5**($[i\ \mathbf{stit}]$) and **S5**($[Agt\ \mathbf{stit}]$) correspond to the fact that $\equiv$, $\mathcal{C}_i$ and $\mathcal{C}_{Agt}$ are equivalence relations, respectively. **KD(X)** corresponds to the fact that $\rightarrow$ is a serial relation, while **(Alt$_\mathbf{X}$)** to the fact that $\rightarrow$ is deterministic. **K(Y)** together with **(Conv$_{\mathbf{X},\mathbf{Y}}$)** and **(Conv$_{\mathbf{Y},\mathbf{X}}$)** correspond to the fact that $\leftarrow$ is the inverse relation of $\rightarrow$ and **(Alt$_\mathbf{Y}$)** to the fact that $\leftarrow$ is deterministic. Finally, **(Rel$_{\Box,[i\ \mathbf{stit}]}$)**, **(AIA)**, **(Rel$_{[i\ \mathbf{rstit}],[i\ \mathbf{stit}]}$)**, **(RatCh)**, **(OneRat)**, **(Rel$_{[i\ \mathbf{stit}],[Agt\ \mathbf{stit}]}$)** and **(NCUH)**, correspond to the Constraints C2, C3, C7, C8, C9, C5* and C6, respectively.

Moreover, it is routine, too, to check that all principles given in Figure 3 are in the so-called Sahlqvist class. This means that they are complete with respect to the defined model classes, cf. [3, Th. 2.42]. $\square$

6. FORMALIZATION OF INFLUENCE

We can now get back to the main issue of the paper, namely the problem of modeling the concept of social influence in STIT. Let us consider the following definition of social influence:

> An agent i influences another agent j to perform a certain (voluntary) action if and only if, i sees to it that j will perform the action.

This definition of social influence is problematic for two reasons: (i) if an agent i sees to it that some state of affairs φ will be true, then φ will *necessarily* be true after i's choice, and (ii) necessity is incompatible with action. Indeed, as emphasized in Section 2, the performance of an action by agent j producing the state of affairs φ, presupposes that this action could have not taken place and that φ possibly would not hold, which means that the action of agent j was not necessary. This point is made clear by the following validity of our logic DR-STIT, that is also a validity of STIT in general. For all $i, j \in Agt$, we have:

$$\models_{\text{DR-STIT}} \neg[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi.$$

PROOF. Let us provide the Hilbert-style proof of this DR-STIT validity by means of the proof calculus given in Section 5:

1. $\vdash$ $[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi \leftrightarrow [i\ \mathbf{stit}]\mathbf{X}([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi)$
2. $\vdash$ $[i\ \mathbf{stit}]\mathbf{X}([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi) \rightarrow [Agt\ \mathbf{stit}]\mathbf{X}([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi)$
   By Axiom **Rel**$_{[i\ \mathbf{stit}],[Agt\ \mathbf{stit}]}$
3. $\vdash$ $[Agt\ \mathbf{stit}]\mathbf{X}([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi) \rightarrow \mathbf{X}\Box([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi)$
   By Axiom **NCUH**
4. $\vdash$ $\mathbf{X}\Box([j\ \mathbf{stit}]\varphi \wedge \Diamond\neg\varphi) \rightarrow \mathbf{X}\Box(\varphi \wedge \Diamond\neg\varphi)$
   By Axiom T for $[j\ \mathbf{stit}]$, Axiom K and necessitation for $\mathbf{X}$ and $\Box$
5. $\vdash$ $\mathbf{X}\Box(\varphi \wedge \Diamond\neg\varphi) \rightarrow \mathbf{X}(\Box\varphi \wedge \Box\Diamond\neg\varphi)$
   By Axiom K for $\Box$, and Axiom K and necessitation for $\mathbf{X}$
6. $\vdash$ $\mathbf{X}(\Box\varphi \wedge \Box\Diamond\neg\varphi) \rightarrow \mathbf{X}(\Box\varphi \wedge \Diamond\neg\varphi)$
   By Axiom T for $\Box$, and Axiom K and necessitation for $\mathbf{X}$
7. $\vdash$ $\mathbf{X}(\Box\varphi \wedge \Diamond\neg\varphi) \rightarrow \bot$
   By Axiom D for $\mathbf{X}$
8. $\vdash$ $\neg[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$
   From 1-7

$\square$

This means that agent i cannot see to it that in the next world agent j deliberatively sees to it that some state of affairs φ is true. As a side note, we observe that $\neg[i\ \mathbf{dstit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$ is valid as well because $[i\ \mathbf{dstit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$ implies $[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$.[4]

However, a non-problematic notion of social influence can be expressed in our logic DR-STIT by means of the special operators $[i\ \mathbf{rdstit}]$. Indeed, these operators allow us to formally represent the idea we have discussed in Section 3, namely that the influencer induces the influencee to perform a certain action by constraining her choice set in such a way that the choice that the influencer wants to be chosen is exactly the one that the influencee would choose, given her preferences. Specifically, we shall say that an agent i influences another agent j to make φ true, denoted by $[i\ \mathbf{infl}\ j]\varphi$, if and only if i sees to it that if agent i's current choice is rational then i is going to deliberatively see to it that φ. That is, for all $i, j \in Agt$ such that $i \neq j$, we define:

$$[i\ \mathbf{infl}\ j]\varphi \stackrel{\text{def}}{=} [i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{rdstit}]\varphi.$$

(We use the operator $[j\ \mathbf{rdstit}]$ rather than $[j\ \mathbf{rstit}]$ since, as emphasized in Section 2, the deliberative STIT is more appropriate than the Chellas STIT to describe the consequences of an agent's voluntary action.) Note that, differently from the formula $[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$, the formula $[i\ \mathbf{infl}\ j]\varphi$ is satisfiable. In order to illustrate this, let us go back to the example of Figure 2 in Section 3. Since agent 2 prefers eating bananas to pears, her only rational choice at moment m_2 is $\{w_7\}$. That is, we assume that $\{w_7\} \in \mathbf{RC}_2$ while $\{w_8\} \notin \mathbf{RC}_2$. From this assumption, it follows that formula $[1\ \mathbf{infl}\ 2]ba$ is true at world w_1. Indeed, $[1\ \mathbf{stit}]\mathbf{X}[2\ \mathbf{rdstit}]ba$ is clearly true at w_1 because for all $v \in \mathcal{C}_1 \circ \twoheadrightarrow (w_1) = \{w_7, w_8\}$ we have: (i) if $\mathcal{C}_2(v) \in \mathbf{RC}_2$ then $M, u \models ba$ for all $u \in \mathcal{C}_2(v)$, and (ii) $M, u \models \neg ba$ for some $u \in \equiv(v)$.

The following proposition highlights some interesting properties of the influence operator $[i\ \mathbf{infl}\ j]$.

PROPOSITION 4. *For all $i, j, k \in Agt$ such that $i \neq j$, $i \neq k$ and $j \neq k$ we have:*

$$\models_{\text{DR-STIT}} [i\ \mathbf{infl}\ j]\varphi \rightarrow [i\ \mathbf{stit}]\mathbf{X}\Diamond\neg\varphi \quad (1)$$

$$\models_{\text{DR-STIT}} ([i\ \mathbf{infl}\ j]\varphi \wedge [i\ \mathbf{infl}\ j]\psi) \rightarrow [i\ \mathbf{infl}\ j](\varphi \wedge \psi) \quad (2)$$

$$\models_{\text{DR-STIT}} \neg[i\ \mathbf{infl}\ j]\top \quad (3)$$

$$\models_{\text{DR-STIT}} \neg[i\ \mathbf{infl}\ j]\bot \quad (4)$$

$$\models_{\text{DR-STIT}} \neg([i\ \mathbf{infl}\ j]\varphi \wedge [i\ \mathbf{infl}\ k]\neg\varphi) \quad (5)$$

$$\models_{\text{DR-STIT}} [i\ \mathbf{infl}\ j][j\ \mathbf{infl}\ k]\varphi \leftrightarrow [i\ \mathbf{infl}\ j]\mathbf{X}[k\ \mathbf{rdstit}]\varphi \quad (6)$$

Validity 1 captures the idea that agent i influences agent j to perform a certain voluntary action only if the result of j's action *is not necessary*. Indeed, as emphasized above, action is incompatible with necessity. Validity (2) characterizes the behavior of the operator $[i\ \mathbf{infl}\ j]$ with conjunction. Note that its converse (i.e., $[i\ \mathbf{infl}\ j](\varphi \wedge \psi) \rightarrow ([i\ \mathbf{infl}\ j]\varphi \wedge [i\ \mathbf{infl}\ j]\psi)$) is not DR-STIT valid. Indeed, the fact that, after i's action, j has a choice available which could possibly make $\varphi \wedge \psi$ false, does not imply that after i's action, j has a choice available which could possibly make φ false and a choice available which could possibly make φ false. Validities (3) and (4) just say that an agent cannot influence another agent to bring about tautologies or contradictions. These two validities follow from Axiom **(OneRat)**. Indeed, Axiom **(OneRat)** guarantees that, after i's action, j has at least one rational choice. Since, it is never the case that an agent deliberatively brings about tautologies or contradictions (i.e., $\neg[i\ \mathbf{dstit}]\top$ and $\neg[i\ \mathbf{dstit}]\bot$ are both valid formulae), it follows that i cannot influence j to bring about tautologies or contradictions. According to validity (5), an agent cannot influence two different agents to bring about conflicting results. Finally, validity (6) provides a characterization of chain of influences: the fact 'i influences j to influence k to make φ true' just means that i influences j to ensure that in the next state if k's current choice is rational, then k deliberatively sees to it that φ.

The operator $[i\ \mathbf{infl}\ j]$ captures the minimal condition for agent i to influence agent j to perform a certain action, namely the fact that i's choice is a sufficient condition for j's action to occur. A stronger notion of influence also requires that j's action would have not occurred had i made a different choice. This stronger notion of influence is captured by the following abbreviation, with $i \neq j$:

$$[i\ \mathbf{sinfl}\ j]\varphi \stackrel{\text{def}}{=} [i\ \mathbf{infl}\ j]\varphi \wedge \neg\Box\mathbf{X}[j\ \mathbf{rstit}]\varphi.$$

where $[i\ \mathbf{sinfl}\ j]\varphi$ has to be read "agent i *strongly* influences agent j to make φ true". The condition $\neg\Box\mathbf{X}[j\ \mathbf{rstit}]\varphi$ guarantees that agent i does not *strongly* influence agent j to make φ true, when j's action of bringing about φ is inevitable, in the sense that, *necessarily*, if in the next state j makes a rational choice then j will bring about φ. It is worth noting that the six validities in the preceding Proposition 4 are preserved by replacing the influence operator $[i\ \mathbf{infl}\ j]$ with the strong influence operator $[i\ \mathbf{sinfl}\ j]$.

7. RELATED WORK

The concept of influence has been modeled by Ingmar Pörn [18, 19] whose logic of action builds upon [11]. However, in this weaker logic, contrary to STIT, it is not contradictory to affirm that an agent i brings it about that another agent j brings it about that φ (i.e., $\mathbf{E}_i\mathbf{E}_j\varphi$ is consistent). The same holds in the definition of the 'bringing it about' operator $\mathbf{E}_i$ proposed by [7], which was extended in [21] to distinctively address the production of outcomes by influencing others. As we observed above, it seems to us that STIT's inconsistency of i's seeing to it that j deliberatively sees to it that φ ($[i\ \mathbf{stit}]\mathbf{X}[j\ \mathbf{dstit}]\varphi$) correctly reflects the so-called negative condition of agency, namely, the fact that an outcome can properly be attributed to an action only when the outcome might not have obtained, had this action not taken place (a different understanding of such a negative condition, however, is assumed by [7]). Moreover neither formalization of the 'bringing it about' operator $\mathbf{E}_i$ includes a way to deal with time, and aspect that is essential, we believe, for capturing how the influencer's action constrains the subsequent behavior of the influencee.

Our idea of the rational choice of an agent, while formally similar to the idea of an obligatory choice in Horty's semantics for deontic logic (see [9, Chapter 4]), has a completely different purpose, being complementary, rather than alternative to a deontic model. In particular, Horty's utilitarian semantics provides a foundation for obligations governing a community of agents, by assigning to each history a single social utility. This social utility determines individual obligations, under the assumption that individuals ought to choose histories having the highest social utility (this is the concept of 'ought' of utilitarian morality), or rather to make a choice that is not dominated by some other choice, according to the social utility of the histories it includes. We, on the contrary, through our notion of a rational choice only want to model how each influencee would act in the context resulting from the influencer's action, if she were to act rationally, where by 'rationally' we only mean, 'according to her individual preferences over the set of choices that are available to her'. Thus, we must allow in principle for as many preference

[4]Note that the formulae $\neg[i\ \mathbf{stit}][j\ \mathbf{dstit}]\varphi$ and $\neg[i\ \mathbf{dstit}][j\ \mathbf{dstit}]\varphi$ are valid as well, when $i \neq j$. These two validities are consequences of the property of independence of choices (Constraint C3 in Definition 1).

orderings over possible choices as there are individuals, and distinguish the notion of rational action from the idea of a morally (or legally) obligatory action. In this way that we can cover both inducements to behave in a socially beneficial ways (e.g. through sanctions or incentives), and inducements to behave antisocially (e.g., through threats or bribes). By combining our logic of influence with a deontic logic, we can then distinguish deontically permissible and impermissible influence patterns, namely cases when the influencer or the influencee violate deontic constraints in exercising the influence or in conforming to it.

8. CONCLUSION

Let's take stock. We have started the paper by raising the challenge of modeling the concept of social influence in STIT theory. Then, we have proposed a variant of STIT with special operators describing the consequences of an agent's rational choice and shown that our logic offers a suitable framework for modeling this concept. On the technical side, we have provided a proof calculus for this logic.

Directions of future work are manifold. An important issue that has not been addressed in this paper is the relationship between the concept of rational (or preferred) choice and the concept of preference over outcomes. Indeed, the fact that the choice of an agent is considered to be rational (or preferred) depends on the fact that, by making this choice, the agent will maximize her preferences over the outcomes. This is one of the fundamental aspect of classical decision theory. The logic DR-STIT, as it stands, has nothing to say about this relationship. In order to overcome this limitation of the logic DR-STIT, we plan to extend it by modal operators for preference such as the ones studied by [23].

Another interesting direction of future research is an extension of our analysis of social influence to the *achievement* STIT operator of [1]. The interesting aspect of this operator is that it allows for a fine-grained characterization of the *counterfactual* dimension of causality in action. Specifically, the achievement STIT operator is a 'backward-looking' operator of agency. That is, in order to say that agent i is the cause of φ (in the achievement STIT sense), one must look in the past and check whether agent i had the possibility of making a different choice resulting in φ to be false now. We believe that the achievement STIT operator as defined by Belnap & Perloff (or, at least, an approximation of it) can be expressed in our logic DR-STIT, by combining in the appropriate way the Chellas STIT operator $[i\ \mathbf{stit}]$, the operator of historical necessity $\Box$, and the 'forward-looking' and 'backward-looking' temporal operators $\mathbf{X}$ and $\mathbf{Y}$.

On the technical side, we plan to look at the computational properties of DR-STIT starting with decidability of the satisfiability problem and then moving to the analysis of the computational complexities of both model checking and the satisfiability problem.

9. ACKNOWLEDGEMENTS

Emiliano Lorini acknowledges the support of the French ANR project EmoTES "Emotions in strategic interaction: theory, experiments, logical and computational studies".

10. REFERENCES

[1] N. Belnap and M. Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54:175–199, 1988.

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

[3] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

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

[5] C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103(1-2):157–182, 1998.

[6] B. J. Chellas. Time and modality in the logic of agency. *Studia Logica*, 51:485–517, 1992.

[7] D. Elgesem. The modal logic of agency. *Nordic Journal of Philosophical Logic*, 2:1–46, 1997.

[8] A. Herzig and E. Lorini. A dynamic logic of agency I: STIT, abilities and powers. *Journal of Logic, Language and Information*, 19(1):89–121, 2010.

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

[10] J. F. Horty and N. Belnap. The deliberative STIT: A study of action, omission, and obligation. *Journal of Philosophical Logic*, 24(6):583–644, 1995.

[11] S. Kanger. Law and logic. *Theoria*, 38:105–132, 1972.

[12] G. W. Leibniz. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*. Open Court, La Salle, Ill., [1719] 1985. Transl. E.M. Huggard.

[13] C. List and W. Rabinowicz. Two intuitions about free will: Alternative possibilities and endorsement. Technical report, London School of Economics, London, 2013.

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

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

[16] G. E. Moore. *Ethics: The Nature of Moral Philosophy*. Oxford University Press, Oxford, [1912] 2005.

[17] P. Panzarasa, N. Jennings, and T. J. Norman. Formalising collaborative decision making and practical reasoning in multi-agent systems. *Journal of Logic and Computation*, 12(1):55–117, 2002.

[18] I. Pörn. *The Logic of Power*. Blackwell, Oxford, 1970.

[19] I. Pörn. On the nature of social order. In J.E. Fenstad, I.T. Frolov, and R. Hilpinen, editors, *Logic, Metodology and Philosophy of Science. Vol. 8*, pages 553–67. North Holland, Amsterdam, 1989.

[20] A. Prior. *Past, Present, and Future*. Clarendon Press, Oxford, 1967.

[21] F. Santos, A. Jones, and J. Carmo. Action concepts for describing organised interaction. In *Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences*, pages 373–382. IEEE Computer Society, 1997.

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

[23] J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.

[24] S. Wölf. Propositional Q-logic. *Journal of Philosophical Logic*, 31:387–414, 2002.

[25] M. Xu. Axioms for deliberative STIT. *Journal of Philosophical Logic*, 27:505–552, 1998.

[26] A. Zanardo. Branching-time logic with quantification over branches: The point of view of modal logic. *Journal of Symbolic Logic*, 61(1):143–166, 1996.