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
1.2.4 [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.

Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

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 \( \varphi \) 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 \( \varphi \) 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 BT+AC structures: branching-time structures (BT) augmented by agent choice functions (AC). A BT structure is made of a set of moments and a tree-like ordering over them. An 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 BT+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 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].

An agent $i$ Chellas-sees-to-it that $\varphi$, denoted by formula $[i \text{ stit}_1] \varphi$, 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 $\varphi$ 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 world $w_3$.

Deliberative STIT satisfies the same positive condition as Chellas STIT plus a negative condition: an agent $i$ deliberatively-sees-to-it that $\varphi$, denoted by formula $[i \text{ dstit}_1] \varphi$, at a certain world $w$ if and only if: (i) agent $i$ Chellas-sees-to-it that $\varphi$ at $w$, and (ii) there exists a world $v$ such that $w$ and $v$ belong to the same moment and $\varphi$ 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 $[i \text{ stit}_1] \varphi$ only requires that $i$’s choice ensures that $\varphi$, the truth of $[i \text{ dstit}_1] \varphi$ also requires that $i$ had the opportunity of making an alternative choice that would not guarantee that $\varphi$ 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 $[i \text{ stit}] \varphi$ at $w$ is consistent with (and is indeed entailed by) the necessity of $\varphi$ at $w$, $[i \text{ dstit}] \varphi$ at $w$ is incompatible with the necessity of $\varphi$ 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.\footnote{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.}

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 deterministic 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 choices.\footnote{We shall not consider achievement STIT of [1].}
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: \textit{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 temporal 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 \(Atrm = \{p, q, \ldots\}\) and a finite set of agents \(Agt = \{1, \ldots, n\}\).

The language \(L_{DR-STIT}(Atrm, Agt)\) of the logic DR-STIT is the set of formulae defined by the following BNF:

\[
\phi ::= p \mid \neg \phi \mid \phi \land \phi \mid \Box \phi \mid X \phi \mid Y \phi \mid [i \ \text{st}] \phi \mid [Agt \ \text{st}] \phi \mid [i \ \text{rst}] \phi
\]

where \(p\) ranges over \(Atrm\) and \(i\) ranges over \(Agt\). The other Boolean constructions \(\lor, \land, \lor, \rightarrow\) and \(\leftrightarrow\) are defined from \(\land\) and \(\lor\) in the standard way.

Operators of the form \([i \ \text{st}]\) are Chellas STIT operators that have been informally discussed in Section 2. Formula \([i \ \text{st}] \phi\) captures the fact that \(\phi\) is guaranteed by a present action of agent \(i\), and has to be read ‘agent \(i\) sees to it that \(\phi\) regardless of what the other agents do’. We shorten the reading of \([i \ \text{st}] \phi\) to ‘agent \(i\) sees to it that \(\phi\)’. \([Agt \ \text{st}]\) is a group STIT operator which captures the fact that \(\phi\) is guaranteed by a present choice of all agents, and has to be read ‘all agents see to it that \(\phi\) by acting together’. The modal operator \([Agt \ \text{st}]\) 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 unindivisible histories [2, Chap. 7]. The dual operators of \([i \ \text{st}]\) and \([Agt \ \text{st}]\) are defined in the usual way:

\[
(i \ \text{st}) \phi \equiv \neg [i \ \text{st}] \neg \phi,
\]

\[
(Agt \ \text{st}) \phi \equiv \neg [Agt \ \text{st}] \neg \phi.
\]

Operators of the form \([i \ \text{rst}]\) 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 \([i \ \text{rst}] \phi\) 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 \(\phi\)” or, “if agent \(i\)’s current choice is a preferred choice of \(i\), then \(i\) sees to it that \(\phi\).”

The dual of the operator \([i \ \text{rst}]\) is defined as follows: \([i \ \text{rst}] \phi \equiv \neg [i \ \text{rst}] \neg \phi\). Note that \([i \ \text{rst}] \top\) 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 \([i \ \text{rst}] \phi \land [i \ \text{rst}] \top\), which is logically equivalent to \([i \ \text{rst}] \phi \land [i \ \text{rst}] \top\), has to be read either “agent \(i\) sees to it that \(\phi\) as a result of her rational choice” or, “agent \(i\) sees to it that \(\phi\) as a result of her preferred choice”.

\(\Box \phi\) stands for ‘\(\phi\) is true regardless of what every agent does’ or ‘\(\phi\) is true no matter what the agents do’ or simply ‘\(\phi\) is necessarily true’. We define the dual of \(\Box\) as follows: \(\Box \phi \equiv \neg \Box \neg \phi\). Note that the operators \([i \ \text{st}]\) and \(\Box\) can be combined in order define the deliberative STIT operator \([i \ \text{dst}]\) we have discussed in Section 2:

\([i \ \text{dst}] \phi \equiv [i \ \text{st}] \phi \land \neg \Box \phi\).

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

\([i \ \text{rdst}] \phi \equiv [i \ \text{rst}] \top \rightarrow [i \ \text{dst}] \phi\).

\([i \ \text{rdst}] \phi\) 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 \(\phi\)” or, “if agent \(i\)’s current choice is a preferred choice of \(i\), then \(i\) deliberatively sees to it that \(\phi\)”.

Finally, \(X\) and \(Y\) are the standard operators ‘next’ (tomorrow) and ‘previous’ (yesterday) of linear temporal logic. \(X \phi\) and \(Y \phi\) has to be read respectively “\(\phi\) is going to be true in the next world” and ‘\(\phi\) 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 \(R\) on \(W\) and an element \(w\) in \(W\), let \(R(w) = \{v \in W | wRv\}\). Moreover, given two binary relations \(R_1\) and \(R_2\) on \(W\) let \(R_1 \circ R_2\) be

\[\text{Our notion of "doing something as the result of a rational (or preferred) choice" is synonym of List & Rabinowicz’s notion of "doing an action with endorsement" [13].}\]
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 agent $Agt$.

**Definition 1.** A Kripke STIT model with discrete time and rational choices is a tuple $M = (\{W, \equiv, \rightarrow\}, \{C_i\}_{i \in Agt}, \{RC_i\}_{i \in Agt}, V)$ where:

- $W$ is a nonempty set of possible worlds;
- $\equiv$ is an equivalence relation on $W$;
- $\rightarrow$ is a serial and deterministic relation on $W$;
- $\leftarrow$ is the inverse relation of $\rightarrow$ (i.e., $\leftarrow = \{(w, v) | v \rightarrow w\}$) and is supposed to be deterministic;
- $C_{Agt}$ and every $C_i$ are equivalence relations on $W$;
- $\forall RC_i$ is a subset of the partition of $W$ induced by the equivalence relation $C_i$;
- $V : Atm \rightarrow 2^W$ is a valuation function for atomic propositions;

and that satisfies the following six constraints:

- **(C1)** for all $w \in W$: if $w.F_v$ then $w \neq v$, with $F$ denoting the transitive closure of the binary relation $\rightarrow$;
- **(C2)** for all $i \in Agt$: $C_i \subseteq \equiv$;
- **(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 $\bigcap_{1 \leq i \leq n} C_i(u_i) \neq \emptyset$;
- **(C4)** for all $w \in W$ and for all $i \in Agt$: there exists $v \in W$ such that $w \equiv v$ and $C_i(v) \in RC_i$;
- **(C5)** for all $w \in W$: $C_{Agt}(w) = \bigcap_{i \in Agt} C_i(w)$;
- **(C6)** $\rightarrow \equiv \subseteq C_{Agt} \circ \leftarrow$.

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

$\equiv$ 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 $m, m', \ldots$.

$\rightarrow$ is the set of direct temporal successors of world $w$, that is to say, $w \rightarrow v$ 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 $\rightarrow$ is serial and deterministic means that every world has exactly one direct temporal successor. $\leftarrow (w)$ defines the set of direct temporal predecessors of world $w$, that is to say, $v \leftarrow w$ 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 $\leftarrow$ is deterministic means that every world has at most one direct temporal predecessor. We do not assume $\leftarrow$ to be serial because past is not necessarily less.

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 $F(v)$ 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 $\rightarrow$, $\leftarrow$ and $F$ are all irreflexive.

Let $T(w) = P(w) \cup \{w\} \cup F(w)$ be the set of worlds that are temporally related with world $w$, where $P = \{(w, v) | v.Fw\}$ is the inverse of the relation $F$ and $P(w)$ is the set of worlds that are in the past of $w$. The fact that the relation $F$ is irreflexive and transitive ensures that $F$ is a strict linear (or total) order on the set $T(w)$. For every world $w$ in $W$, we call the linearly ordered set $(T(w), F)$ the history going through $w$. Note that, because of the seriality of the relation $\rightarrow$, 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 $C_i(w)$ 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 $\{h \in Hist | \exists w \in C_i(w) \text{ such that } v \in h\}$. 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 $C_i$ 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 $C_1(u_1)$ is a possible choice for agent 1 at $w$, $C_2(u_2)$ is a possible choice for agent 2 at $w$, then $C_1(u_1) \cap C_2(u_2)$ is a possible choice for agent 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 $C_{Agt}$ denote the partition of the set of worlds $W$ induced by the equivalence relation $C_{Agt}$. This partition characterizes agent $i$’s set of choices in the model $M$. The set $RC_i \subseteq C_i$ characterizes
implies that if an agent can choose between two histories at a later
each agent does not have a choice between these two histories. This
histories come together in some future moment then, in the present,
ries
a given Kripke STIT model with discrete time and rational choices
no choice between undivided histo-
corresponds to the property of
all agents at
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 mo-
ment, 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 undisjoined histo-
gies 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 \( \varphi \) of the logic DR-STIT is evaluated with respect to
a given Kripke STIT model with discrete time and rational choices
\( M = (W, \equiv, \rightarrow, \varnothing, \{C_i\}_{i \in \text{Agt}}, \mathcal{C}_{\text{Agt}}, \{R_{C_i}\}_{i \in \text{Agt}}, \mathcal{V}) \) and a world
\( w \) in \( M \). We write \( M, w \models \varphi \) to mean that \( \varphi \) is true at world \( w \) in
\( M \). The truth conditions of DR-STIT formulate are then defined as follows:

\[
\begin{align*}
M, w \models \rho & \iff \text{\( w \in V(p) \)} \\
M, w \models \neg \varphi & \iff M, w \not\models \varphi \\
M, w \models \varphi \land \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 X \varphi & \iff \forall v \in \rightarrow(w) : M, v \models \varphi \\
M, w \models Y \varphi & \iff \forall v \in \leftarrow(w) : M, v \models \varphi \\
M, w \models [i \text{ stit}] \varphi & \iff \forall v \in C_i(w) : M, v \models \varphi \\
M, w \models [i \text{ rstIt}] \varphi & \iff \forall v \in \{C_i(w) \in R_{C_i} : M, v \models \varphi \}
\end{align*}
\]

For any formula \( \varphi \) of the language \( \mathcal{L}_{\text{DR-STIT}}(\mathcal{A}_{\text{Atm}}, \text{Agt}) \), we write
\( \models_{\text{DR-STIT}} \varphi \) if \( \varphi \) is DR-STIT valid, i.e., for all Kripke STIT models
with discrete time and rational choices and M for all worlds \( w \)
in \( M \), we have \( M, w \models \varphi \). We say that \( \varphi \) 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 proposi-
tional calculus (PC) as well as modus ponens (MP). Moreover,
we have all principles of the normal modal logic S5 for every
operator \([i \text{ stit}]\), for the operator \([\text{Agt stit}]\) and for the operator \( \Box \),
all principles of the normal modal logic KD for the temporal
operator \( X \) and all principles of the normal modal logic K for the
temporal operator \( Y \). That is, we have Axiom K for each operator:
\( \Box \varphi \land \Box (\varphi \rightarrow \psi) \rightarrow \Box \psi \) with \( \Box \in \{\Box, X, Y, [\text{Agt stit}]\} \cup \{[i \text{ stit}]\}_{i \in \text{Agt}} \).
We have Axiom D for the temporal operator \( X \:
\neg(X\varphi \land X\neg\varphi) \). We have Axiom 4 for \( \Box, [\text{Agt stit}] \) and for every
\( [i \text{ stit}] \): \( \Box \varphi \rightarrow [i \text{ stit}] \varphi \) with \( \Box \in \{\Box, [\text{Agt stit}]\} \cup \{[i \text{ stit}]\}_{i \in \text{Agt}} \).
Furthermore, we have Axiom T for \( \Box, [\text{Agt stit}] \) and for every
\( [i \text{ stit}] \): \( \Box \varphi \rightarrow \varphi \) with \( \Box \in \{\Box, [\text{Agt stit}]\} \cup \{[i \text{ stit}]\}_{i \in \text{Agt}} \).
We have Axiom B for \( \Box, [\text{Agt stit}] \) and for every \( [i \text{ stit}] \):
\( \varphi \rightarrow [i \text{ stit}] \Box \varphi \) with \( \Box \in \{\Box, [\text{Agt stit}]\} \cup \{[i \text{ stit}]\}_{i \in \text{Agt}} \).
Finally we have the rule of necessitation for each modal operator:
\( \Box \varphi \rightarrow [i \text{ stit}] \varphi \) with \( \Box \in \{\Box, [\text{Agt stit}]\}, X, Y \} \cup \{[i \text{ stit}]\}_{i \in \text{Agt}} \).

We have principles for the temporal operators and for the
relationship between time and action. (AltX) and (AltY) are the basic
axioms for the determinism of 'tomorrow' and 'yesterday'.
(ConvX, Y) and (ConvX, Y) are the basic interaction axioms
between 'tomorrow' and 'yesterday' according to which if \( \varphi \) is true
in the present, then tomorrow is going to be true that yesterday \( \varphi \) has
been true" and "if \( \varphi \) is true in the present, then yesterday has
been true that tomorrow \( \varphi \) is going to be true".

(ReL[i, stit]) and (AIA) are the two central principles in Xu's
axiomatization of the Chellas STIT operators \( [i \text{ stit}] \) [25].
According to Axiom (ReL[i, stit]), if \( \varphi \) is true regardless of what ev-
ery agent does, then every agent sees to it that \( \varphi \). In other words,
an agent brings about those facts that are inevitable.According to
Axiom (ReL[i stit], [Agt 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 undisjoined histories' (Constraint C6 in Definition
1): if in the next world \( \varphi \) is going to be possible then the actual
collective choice of all agents will possibly result in a world in
which \( \varphi \) is true.

Axiom (ReL[i, rstIt], [i stit]) is the basic interaction principle
between the rational choice operator \([i \text{ rstIt}]\) and the Chellas STIT
operator \([i \text{ stit}]\). 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
(OnRat), an agent always has a rational choice in her repertoire.

Theorem 1. The set of DR-STIT validities is completely axi-
omatized by the principles given in Figure 3.

Proof. (Sketch) Space restrictions prevent us from giving a de-
tailed 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 re-
tect 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 pro-
viding a DR-STIT semantics in terms of standard KDRs (SKDRs for short), that is, tuples of the form \( (W, \equiv, \rightarrow, \varnothing, \{C_i\}_{i \in \text{Agt}}, \mathcal{C}_{\text{Agt}}, \{R_{C_i}\}_{i \in \text{Agt}}, \mathcal{V}) \)
where \( W, \equiv, \rightarrow, \varnothing, \{C_i\}_{i \in \text{Agt}}, \mathcal{C}_{\text{Agt}} \) and \( \mathcal{V} \) are exactly
as in Definition 1 and \( \{R_{C_i}\}_{i \in \text{Agt}} \) is a family of binary rela-
tions on \( W \) that satisfy the following three constraints for all
\( i \in \text{Agt} \):

\[
\begin{align*}
\text{(C7)} & \text{ for all } w \in W : \text{ if } R_{C_i}(w) \neq \emptyset \text{ then } R_{C_i}(w) = C_i(w); \\
\text{(C8)} & \text{ for all } w, v \in W : \text{ if } v \in C_i(w) \text{ and } R_{C_i}(w) \neq \emptyset \text{ then } R_{C_i}(v) \neq \emptyset; \\
\text{(C9)} & \text{ for all } w \in W : \text{ there exists } v \in W \text{ such that } w \equiv v \text{ and } R_{C_i}(v) \neq \emptyset.
\end{align*}
\]

It is easy to prove that:

Lemma 1. For every formula \( \varphi \) in \( \mathcal{L}_{\text{DR-STIT}}(\text{Atm}, \text{Agt}) \), \( \varphi \) is
satisfiable relative to the class of KDRs iff it is satisfiable relative
to the class of SKDRs.
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 $\varphi$ in $\mathcal{L}_{\text{DR-STIT}}(\text{Atm, Agt})$, $\varphi$ 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 $M$. In other words, we show that the direction $C_{\text{Agt}}(w) \subseteq \bigcap_{i \in \text{Agt}} C_i(w)$ 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, $SS(\square)$, $SS([i \text{ stit}])$ and $SS([\text{Agt stit}])$ correspond to the fact that $\exists_i C_i$ and $C_{\text{Agt}}$ are equivalence relations, respectively. $KD(Y)$ corresponds to the fact that $\rightarrow$ is a serial relation, while $KD(X)$ to the fact that $\rightarrow$ is deterministic. $K(Y)$ together with $(\text{Conv}_X, Y)$ and $(\text{Conv}_Y, X)$ correspond to the fact that $\rightarrow$ is the inverse relation of $\rightarrow$ and $(\text{Alt}_X)$ to the fact that is deterministic. Finally, $(\text{Rel}_{[i \text{ stit}]}), (\text{AIA}), (\text{Rel}_{[i \text{ stit}]}), (\text{Rel}_{[\text{Agt stit}]}), (\text{Rel}_{[\text{Agt stit}]}), (\text{Agt})$, $(\text{RC})$, and $(\text{NCUH})$, 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].

6. **Fomalization 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 $\varphi$ will be true, then $\varphi$ 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 $\varphi$, presupposes that this action could have not taken place and that $\varphi$ 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 $i, j \in \text{Agt}$, we have:

$$\vdash_{\text{DR-STIT}} \neg[i \text{ stit}]X[j \text{ dstit}]\varphi.$$
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 $\varphi$ is true. As a side note, we observe that $\neg[i \ dstit \ j \ dstit] \varphi$ is valid as well because $[i \ dstit \ j \ dstit] \varphi$ implies $[i \ stit \ j \ dstit] \varphi$.4

However, a non-problematic notion of social influence can be expressed in our logic DR-STIT by means of the special operators $[i \ rdstit]$. 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 $\varphi$ true, denoted by $[i \ infl \ j] \varphi$, 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 $\varphi$. That is, for all $i, j \in \text{Agt}$ such that $i \neq j$, we define:

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

(We use the operator $[j \ rdstit]$ rather than $[j \ stit]$ 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 $[i \ stit \ X \ j \ dstit] \varphi$, the formula $[i \ infl \ j] \varphi$ 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 $n_2$ is $\{w_7\}$. That is, we assume that $\{w_7\} \in RC_2$ while $\{w_8\} \notin RC_2$. From this assumption, it follows that formula $[1 \ infl \ 2] \varphi$ is true at world $w_1$. Indeed, $[1 \ stit \ X \ 2 \ rdstit] \varphi$ is clearly true at $w_1$ because for all $v \in C_1 \varphi \rightarrow (w_{M \in C_2} (u_1) = (w_7, w_8) \ll X$ we have: $(i) C_2(v) \notin RC_2$ then $M, u \models \varphi$ for all $u \in C_2 (v)$, and $(ii) M, u \models \neg \varphi$ for some $u \in (\equiv \ll v)$. The following proposition highlights some interesting properties of the influence operator $[i \ infl \ j]$.

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

1. $\models \text{DR-STIT} [i \ infl \ j] \varphi \rightarrow [i \ stit \ X \varphi]$  
2. $\models \text{DR-STIT} ([i \ infl \ j] \varphi \land [i \ infl \ j] \psi) \rightarrow [i \ infl \ j] (\varphi \land \psi)$  
3. $\models \text{DR-STIT} \neg [i \ infl \ j] \top$  
4. $\models \text{DR-STIT} \neg [i \ infl \ j] \bot$  
5. $\models \text{DR-STIT} \neg ([i \ infl \ j] \varphi \land [i \ infl \ k] \neg \varphi)$  
6. $\models \text{DR-STIT} [i \ infl \ j] [k \ infl \ k] \varphi \leftrightarrow [i \ infl \ j] [k \ rdstit] \varphi$

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 incompatibile with necessity. Validity 2 characterizes the behavior of the operator $[i \ infl \ j]$ with conjunction. Note that its converse (i.e., $[i \ infl \ j] (\varphi \land \psi) \rightarrow ([i \ infl \ j] \varphi \land [i \ infl \ j] \psi)$) is not DR-STIT valid. Indeed, the fact that, after $i$’s action, $j$ has a choice available which could possibly make $\varphi \land \psi$ false, does not imply that after $i$’s action, $j$ has a choice available which could possibly make $\varphi$ false and a choice available which could possibly make $\psi$ 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., $\neg [i \ dstit] \top$ and $\neg [i \ dstit] \bot$ 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 $\varphi$ 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 $\varphi$. The operator $[i \ infl \ j]$ 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 $i \neq j$:

$$[i \ sinfl \ j] \varphi \overset{\text{def}}{=} [i \ infl \ j] \varphi \land \neg \Box (j \ rdstit) \varphi,$$

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

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 $\varphi$ (i.e., $E_i E_j \varphi$ is consistent). The same holds in the definition of the ‘bringing it about’ operator $E_i$, 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 $\varphi$ ($[i \ stit \ X \ j \ dstit] \varphi$) 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 $E_i$, 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 influencer.

Our idea of the rational choice of an agent, while formally similar to the idea of an obligatory choice in Horthy’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, Horthy’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 influence 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

891
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 $\varphi$ (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 $\varphi$ 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 $[i \triangleleft \varphi]$, the operator of historical necessity $\square$, and the ‘forward-looking’ and ‘backward-looking’ temporal operators $X$ and $Y$.

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