Temporal STIT logic and its application to normative reasoning

Emiliano Lorini

HAL Id: hal-01130572

https://hal.archives-ouvertes.fr/hal-01130572

Submitted on 12 Mar 2015

Temporal STIT logic and its application to normative reasoning

Emiliano Lorini*

IRIT-CNRS, Toulouse, France

I present a variant of STIT with time, called T-STIT (Temporal STIT), interpreted in standard Kripke semantics. On the syntactic level, T-STIT is nothing but the extension of atemporal individual STIT by: (i) the future tense and past tense operators, and (ii) the operator of group agency for the *grand coalition* (the coalition of all agents). A sound and complete axiomatisation for T-STIT is given. Moreover, it is shown that T-STIT supports reasoning about interesting normative concepts such as the concepts of achievement obligation and commitment.

Keywords: STIT logic; temporal logic; axiomatisation; deontic logic

1. Introduction

STIT logic (the logic of *Seeing to it That*; Belnap, Perloff, & Xu, 2001) 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'. Horty (2001) extends the STIT logic of Belnap et al. (2001) with operators of group agency in order to express sentences of the form 'the group of agents C sees to it that φ is true'. Following Lorini and Schwarzentruber (2011), one might use the terms 'individual STIT logic' and 'group STIT logic' to designate respectively the STIT logic of Belnap et al. (2001) (in which only the actions of agents are described) and Horty's (2001) variant of STIT logic (in which both actions of agents and joint actions of groups are represented).

The original semantics for STIT given by Belnap et al. (2001) 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 a certain 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. As shown by Balbiani, Herzig, and Troquard (2008), Lorini and Schwarzentruber (2011), and Herzig and Schwarzentruber (2008), however, both 'atemporal individual STIT' (i.e., individual STIT without tense operators in the object language) axiomatised by Belnap et al. (2001, Chapter 17), and 'atemporal group STIT' (i.e., group STIT without tense operators) can be 'simulated' in standard Kripke semantics. A similar idea is proposed by Kooi and Tamminga (2008), who introduce the concept of 'consequential model'. Consequential models are equivalent to the Kripke atemporal group STIT models used by Herzig and Schwarzentruber (2008) and Lorini and Schwarzentruber (2011), in which the authors abstract away from the branching-time account of STIT.

---

*Email: lorini@irit.fr

The present article goes beyond previous work on atemporal STIT by presenting a variant of STIT logic with time interpreted in standard Kripke semantics and by providing a sound and complete axiomatisation for this logic. I call this variant of STIT logic T-STIT (Temporal STIT). On the syntactic level, the logic T-STIT is nothing but the extension of atemporal individual STIT by: (i) the future tense and past tense operators, and (ii) the operator of group agency in Horty's (2001) sense for the *grand coalition* (the coalition of all agents).

The main motivation of this work is that past research on the mathematical properties of STIT (e.g., completeness and decidability) has mainly focused on atemporal STIT, while extensions of STIT by tense operators are far less well studied and understood. The interest of studying them is that they offer valuable formal languages for representing a variety of normative concepts such as commitment (Bentahar, Moulin, Meyer, & Chaib-draa, 2004; Desai, Narendra, & Singh, 2008; Singh, 2008) and achievement obligation (i.e., the obligation to perform a given action at some point in the future; Broersen, Dastani, & van der Torre, 2003; Governatori & Rotolo, 2010) that have an intrinsic *temporal* nature. These concepts are fundamental for understanding normative relationships between individuals in a society and have become useful abstractions for the design of multi-agent systems since they can be used to model a variety of interactive situations like contracts, agreements, negotiation, dialogue, and argumentation.

The rest of the article is organised as follows. In Section 2, the logic T-STIT is presented and a complete axiomatisation for this logic is given. An application of T-STIT to the formalisation of normative concepts such as achievement obligation and commitment is given in Section 3. It is shown that that T-STIT allows us to capture subtle temporal properties of these normative concepts such as *persistence* (i.e., the conditions under which a commitment persists over time). In Section 4, related work on STIT logic is discussed.

2. A STIT logic with time

This section presents the syntax and a Kripke-style semantics for T-STIT (Subsections 2.1 and 2.2). An axiomatisation of T-STIT is provided in Subsection 2.3, while in Subsection 2.4 it is proved that this axiomatisation is sound and complete with respect to the given semantics.

In the logic T-STIT, the so-called Chellas's STIT operators (Chellas, 1992) are taken as primitive operators of agency. As pointed out by Xu (1998) and Horty and Belnap (1995), so-called deliberative STIT operators and Chellas's STIT operators are interdefinable and just differ in the choice of primitive operators.

2.1. Syntax

Assume a countably infinite set of propositional atoms denoting basic facts $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$.

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

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [i]\varphi \mid [Agt]\varphi \mid \Box\varphi \mid \mathsf{G}\varphi \mid \mathsf{H}\varphi$$

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 $[i]$ are Chellas's STIT operators. The formula $[i]\varphi$ 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'. I shorten the reading of $[i]\varphi$ to 'agent i sees to it that φ'. The crucial aspect of STIT theory is that an agent i's action is described in terms of the result that agent i brings about by her acting. For example, i's action of killing another agent j is described by the fact that i sees to it that j is dead. I define the dual of the operator $[i]$ as follows: $\langle i \rangle \varphi \stackrel{\text{def}}{=} \neg[i]\neg\varphi$.

$[Agt]$ 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 dual of the operator $[Agt]$ is defined as expected: $\langle Agt \rangle \varphi \stackrel{\text{def}}{=} \neg[Agt]\neg\varphi$. The modal operator $[Agt]$ will be fundamental in Section 2.2 in order to axiomatise a basic property relating action and time studied in STIT: the so-called property of *no choice between undivided histories* (Belnap et al., 2001, Chapter 7).

$\Box\varphi$ 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'. I define the dual of $\Box$ as follows: $\Diamond\varphi \stackrel{\text{def}}{=} \neg\Box\neg\varphi$. Note that the operators $[i]$ and $\Diamond$ can be combined in order to express what agents can do: $\Diamond[i]\varphi$ means 'agent i can see to it that φ'. Moreover, the operators $[i]$ and $\Box$ can be combined in order define the deliberative STIT operators $[i \ dstit:]$ studied by Horty and Belnap (1995): $[i \ dstit: \varphi] \stackrel{\text{def}}{=} [i]\varphi \wedge \neg\Box\varphi$.

Finally, G and H are tense operators that are respectively used to express facts that are always true in the *strict* future and facts that are always true in the past. $\mathsf{G}\varphi$ means 'φ will always be true in the future' and $\mathsf{H}\varphi$ means 'φ has always been true in the past'. I define the dual of the future tense operator G as follows: $\mathsf{F}\varphi \stackrel{\text{def}}{=} \neg\mathsf{G}\neg\varphi$. $\mathsf{F}\varphi$ means 'φ will be true at some point in the future'. Moreover, I define the dual of the past tense operator H as follows: $\mathsf{P}\varphi \stackrel{\text{def}}{=} \neg\mathsf{H}\neg\varphi$. $\mathsf{P}\varphi$ means 'φ has been true at some point in the past'.

The following abbreviations will also be convenient:

$$\mathsf{G}^*\varphi \stackrel{\text{def}}{=} \varphi \wedge \mathsf{G}\varphi;$$
$$\mathsf{F}^*\varphi \stackrel{\text{def}}{=} \neg\mathsf{G}^*\neg\varphi.$$

$\mathsf{G}^*\varphi$ stands for 'φ is true in the present and will always be true', whereas $\mathsf{F}^*\varphi$ stands for 'φ is true in the present or will be true at some point in the future'.

2.2. A Kripke semantics for STIT logic with time

The basic notion in the semantics is the notion of a temporal Kripke STIT model that is nothing but a multi-relational Kripke model with special constraints on the accessibility relations. For notational convenience, in what follows I am 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 $\mathcal{R}(w) = \{v \in W | (w, v) \in \mathcal{R}\}$. Moreover, given two binary relations $\mathcal{R}_1$ and $\mathcal{R}_2$ on W let $\mathcal{R}_1 \circ \mathcal{R}_2$ be the standard operation of composition between binary relations. Temporal Kripke STIT models can be seen as extensions of Zanardo's (1996) Ockhamist frames by a choice component, i.e., by accessibility relations for the individual choices of the agents and an accessibility relation for the collective choice of the grand coalition Agt.

Definition 1 (Temporal Kripke STIT model). *The class of temporal Kripke STIT models includes all tuples $M = (W, \mathcal{R}_\Box, \{\mathcal{R}_i | i \in Agt\}, \mathcal{R}_{Agt}, \mathcal{R}_\mathsf{G}, \mathcal{R}_\mathsf{H}, \mathcal{V})$ where:*

- *W is a nonempty set of possible worlds;*
- *$\mathcal{R}_\Box$, every $\mathcal{R}_i$ and $\mathcal{R}_{Agt}$ are equivalence relations between worlds in W such that:*
  (C1) $\mathcal{R}_i \subseteq \mathcal{R}_\Box$;

(C2) *for all $u_1, \ldots, u_n \in W$: if $(u_i, u_j) \in \mathcal{R}_\square$ for all $i, j \in \{1, \ldots, n\}$ then $\bigcap_{1 \leq i \leq n} \mathcal{R}_i(u_i) \neq \emptyset$;*

(C3) *for all $w \in W$: $\mathcal{R}_{Agt}(w) = \bigcap_{i \in Agt} \mathcal{R}_i(w)$;*

- $\mathcal{R}_G$ *and* $\mathcal{R}_H$ *are binary relations between worlds in W such that* $\mathcal{R}_G$ *is serial and transitive,* $\mathcal{R}_H$ *is the inverse relation of* $\mathcal{R}_G$ *(i.e.,* $\mathcal{R}_H = \mathcal{R}_G^{-1} = \{(w, v) | (v, w) \in \mathcal{R}_G\}$*), and:*

(C4) *for all $w, v, u \in W$: if $v, u \in \mathcal{R}_G(w)$ then $u \in \mathcal{R}_G(v)$ or $v \in \mathcal{R}_G(u)$ or $u = v$;*

(C5) *for all $w, v, u \in W$: if $v, u \in \mathcal{R}_H(w)$ then $u \in \mathcal{R}_H(v)$ or $v \in \mathcal{R}_H(u)$ or $u = v$;*

(C6) $\mathcal{R}_G \circ \mathcal{R}_\square \subseteq \mathcal{R}_{Agt} \circ \mathcal{R}_G$;

(C7) *for all $w \in W$: if $v \in \mathcal{R}_\square(w)$ then $v \notin \mathcal{R}_G(w)$;*

- $\mathcal{V} : Atm \longrightarrow 2^W$ *is a valuation function for atomic formulas.*

The valuation function $\mathcal{V}$ is used to identify those states in a model in which a given atomic proposition is true. Specifically, $w \in \mathcal{V}(p)$ means that p is true at world w.

$\mathcal{R}_\square(w)$ is the set of worlds that are alternative to the world w. Following the Ockhamist's view of time (Prior, 1967; Thomason, 1984; Zanardo, 1996), I call the equivalence classes induced by the equivalence relation $\mathcal{R}_\square$ *moments*.[1] The set of all moments in the model M is denoted by *Mom* and the elements in *Mom* are denoted by $m, m', \ldots$.

$\mathcal{R}_G(w)$ defines the set of worlds that are in the *strict* future of world w, where the strict future does not include the present. $\mathcal{R}_H(w)$ defines the set of worlds that are in the *past* of world w. The Constraint C7 ensures that if two worlds belong to the same moment then one of them cannot be in the future of the other. Since the relation $\mathcal{R}_\square$ is reflexive, the Constraint C7 implies the irreflexivity of the relation $\mathcal{R}_G$, i.e., for all $w \in W$ we have $w \notin \mathcal{R}_G(w)$. The fact that the relation $\mathcal{R}_G$ is transitive and irreflexive just means that it is a strict partial order on the set W. The Constraint C4 ensures that time is connected towards the future, while the Constraint C5 ensures that time is connected towards the past.

Let $\mathcal{T}(w) = \mathcal{R}_H(w) \cup \{w\} \cup \mathcal{R}_G(w)$ be the set of worlds that are temporally related with world w. The fact that the relation $\mathcal{R}_G$ is irreflexive and transitive together with the Constraints C4 and C5 ensure that $\mathcal{R}_G$ is a strict linear (or total) order on the set $\mathcal{T}(w)$. For every world w in W, I call the linearly ordered set $(\mathcal{T}(w), \mathcal{R}_G)$ the history going through w. For notational convenience, I write h_w instead of $(\mathcal{T}(w), \mathcal{R}_G)$. Note that, because of the seriality of the relation $\mathcal{R}_G$, every history h_w is infinite.

This highlights that there is a one-to-one correspondence between worlds and histories, as for every world w there exists a unique history going through it. In other words, one can interchangeably use the term 'world' and 'history going through a certain world' without lost of generality. As every world in a model is identified with a unique history going through it, the equivalence relation $\mathcal{R}_\square$ can also be understood as an equivalence relation between historic alternatives: $(w, v) \in \mathcal{R}_\square$ means that the history going through v is alternative to the history going through w.

From the temporal relation $\mathcal{R}_G$ over worlds in W, we can define the following relation $<$ over moments in *Mom*, where $m < m'$ means that moment m' is in the strict future of moment m.

Definition 2 (Ordering of moments). *For all $m, m' \in Mom$: let $m < m'$ if and only if there are $w \in m$ and $v \in m'$ such that $(w, v) \in \mathcal{R}_G$.*

For every world w, the set $\mathcal{R}_i(w)$ identifies agent i's *actual* choice at w, that is to say, the set of all alternatives that is forced by agent i's actual choice at w. Because of the

one-to-one correspondence between worlds and histories, one can also identify agent i's *actual* choice at w with the set of histories $\{h_v : v \in \mathcal{R}_i(w)\}$. In other words, in T-STIT an agent chooses among different sets of histories.

Constraint C1 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 $\mathcal{R}_i$ induces a partition of the set $\mathcal{R}_\square(w)$. An element of this partition is a choice that is *possible (or available)* for agent i at w.

Constraint C2 expresses the so-called assumption of *independence of agents* or *independence of choices*: if $\mathcal{R}_1(u_1)$ is a possible choice for agent 1, $\mathcal{R}_2(u_2)$ is a possible choice for agent 2,..., $\mathcal{R}_n(u_n)$ is a possible choice for agent n, then their intersection is nonempty. More intuitively, this means that agents can never be deprived of choices due to the choices made by other agents.

For every world w, the set $\mathcal{R}_{Agt}(w)$ identifies the *actual* choice of group *Agt* at w – that is to say, the set of all alternatives that is forced by the collective choice of all agents at w. Constraint C3 just says that the set of alternatives that is forced by the collective choice of all agents at w is equal to the pointwise intersection of the sets of alternatives that are forced by the individual choices of the agents in *Agt* at w. In other words, the choice of a group corresponds to the intersection of the choices of the individuals in the group. This corresponds to the notion of joint action proposed by Horty (2001), 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 (Belnap et al., 2001, Chapter 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. The Constraint C6 is crucial in order to prove that the relation $<$ defined above is a tree-like ordering of the moments in *Mom*.

Proposition 3. *The relation $<$ satisfies the following four properties for all $m, m', m'' \in$ Mom:*
(Irreflexivity) $m \not< m$;
(Transitivity) if $m < m'$ and $m' < m''$ then $m < m''$;
(Asymmetry) if $m < m'$ then $m' \not< m$;
(No backward branching) if $m' < m$ and $m'' < m$ then $m' < m''$ or $m'' < m'$ or $m' = m''$.

Proof. The irreflexivity of $<$ follows from the Constraint C7. Indeed, suppose that $m < m$. This implies that there are $w, v \in W$ such that $(w, v) \in \mathcal{R}_\square$ and $(w, v) \in \mathcal{R}_G$. But this is in contradiction with the Constraint C7.

Transitivity of $<$ follows from the transitivity of the relation $\mathcal{R}_G$ and from the Constraints C1, C3 and C6. Suppose that $m < m'$ and $m' < m''$. Therefore, for some arbitrary worlds w_1, w_2, w_3, w_4 we have $w_1 \in m$, $w_2, w_3 \in m'$, $w_4 \in m''$, $w_2 \in \mathcal{R}_G(w_1)$ and $w_4 \in \mathcal{R}_G(w_3)$. By the Constraint C6, it follows that there is $w_5 \in \mathcal{R}_{Agt}(w_1)$ such that $w_3 \in \mathcal{R}_G(w_5)$. The Constraints C1 and C3 together imply that $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$. Hence, there is $w_5 \in m$ such that $w_3 \in \mathcal{R}_G(w_5)$. Thus, by the transitivity of $\mathcal{R}_G$, there is $w_5 \in m$ such that $w_4 \in \mathcal{R}_G(w_5)$. It follows that $m < m''$.

The asymmetry of $<$ follows from its irreflexivity.

No backward branching follows from the Constraints C5 and C6. Suppose that $m' < m$ and $m'' < m$. Therefore, for some arbitrary worlds w_1, w_2, w_3, w_4 we have $w_1, w_2 \in m$, $w_3 \in m'$, $w_4 \in m''$, $w_3 \in \mathcal{R}_H(w_1)$ and $w_4 \in \mathcal{R}_H(w_2)$. By the Constraint C6, it follows that there is $w_5 \in \mathcal{R}_{Agt}(w_3)$ such that $w_5 \in \mathcal{R}_H(w_2)$. Thus, since $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$, there is $w_5 \in m'$ such that $w_5 \in \mathcal{R}_H(w_2)$. Moreover, by the Constraint C5 and the fact that $w_4 \in \mathcal{R}_H(w_2)$, it follows that there is $w_5 \in m'$ such that $w_5 \in \mathcal{R}_H(w_4)$ or $w_4 \in \mathcal{R}_H(w_5)$ or $w_4 = w_5$. Since $w_4 \in m''$, the latter implies that $m' < m''$ or $m'' < m'$ or $m' = m''$. $\blacksquare$

Another interesting property of temporal Kripke STIT models that follows from the Constraints C6 and C7 is the so-called property of *past isomorphism* (PI), which is similar to the property of *past isomorphism* of Zanardo's (1996) Ockhamist frames.

Proposition 4. *For all $w, v \in W$, if $(w, v) \in \mathcal{R}_\square$ then there exists an order-isomorphism f between $\mathcal{R}_H(w)$ and $\mathcal{R}_H(v)$ such that, for all $u \in \mathcal{R}_H(w)$, $(u, f(u)) \in \mathcal{R}_{Agt}$.*[2]

Proof. First of all note that, by Constraints C1 and C3, $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$.

Assume $(w, v) \in \mathcal{R}_\square$. By Constraint C7 and the fact that $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$, every world has at most one $\mathcal{R}_{Agt}$-equivalent in any history and, hence, by Constraint C6, every world u such that $u \in \mathcal{R}_H(w)$ has exactly one $\mathcal{R}_{Agt}$-equivalent in $\mathcal{R}_H(v)$. Therefore, the restriction of $\mathcal{R}_{Agt}$ to the set $\mathcal{R}_H(w) \times \mathcal{R}_H(v)$ is an order-preserving bijective function. $\blacksquare$

Given a temporal Kripke STIT model $M = (W, \mathcal{R}_\square, \{\mathcal{R}_i | i \in Agt\}, \mathcal{R}_{Agt}, \mathcal{R}_G, \mathcal{R}_H, \mathcal{V})$, a world w and a formula φ, I write $M, w \models \varphi$ to mean that φ is true at world w in M. The truth conditions of formulas are then defined as follows:

$$M, w \models p \Longleftrightarrow w \in \mathcal{V}(p)$$
$$M, w \models \neg\varphi \Longleftrightarrow M, w \not\models \varphi$$
$$M, w \models \varphi \wedge \psi \Longleftrightarrow M, w \models \varphi \text{ and } M, w \models \psi$$
$$M, w \models \square\varphi \Longleftrightarrow \forall v \in \mathcal{R}_\square(w) : M, v \models \varphi$$
$$M, w \models [i]\varphi \Longleftrightarrow \forall v \in \mathcal{R}_i(w) : M, v \models \varphi$$
$$M, w \models [Agt]\varphi \Longleftrightarrow \forall v \in \mathcal{R}_{Agt}(w) : M, v \models \varphi$$
$$M, w \models \mathsf{G}\varphi \Longleftrightarrow \forall v \in \mathcal{R}_\mathsf{G}(w) : M, v \models \varphi$$
$$M, w \models \mathsf{H}\varphi \Longleftrightarrow \forall v \in \mathcal{R}_\mathsf{H}(w) : M, v \models \varphi$$

Example 5. *Figure 1 provides an example that clearly illustrates the semantics of T-STIT. At world w in the temporal Kripke STIT model M represented in the figure, agent 2 sees to it that p is true (i.e., $M, w \models [2]p$). Indeed, p holds at every world in agent 2's choice at w. Moreover, at w agent 1 sees to it that p or q is true (i.e., $M, w \models [1](p \vee q)$) because either p or q hold at every world in agent 1's choice at w. Finally, at w the group $\{1, 2\}$ sees to it that q will be true at some point in the future (i.e., $M, w \models [\{1, 2\}]\mathsf{F}q$) because q holds at some future point of every history in the group $\{1, 2\}$'s choice at w.*

Given a T-STIT formula φ, I say that φ is T-STIT *valid*, denoted by $\models_{\text{T-STIT}} \varphi$, if and only if for every temporal Kripke STIT model M and for every world w in M we have $M, w \models \varphi$. I say that φ is satisfiable in T-STIT if and only if $\neg\varphi$ is not T-STIT valid.

The following proposition provides an example of interesting T-STIT validities.

Proposition 6. *The following two formulas are T-STIT valid:*

$$\mathsf{G}\Diamond\mathsf{G}^*\varphi \rightarrow \langle Agt\rangle\mathsf{G}\varphi$$
$$\mathsf{G}\Diamond(\mathsf{G}^*\varphi \wedge \mathsf{F}^*\psi) \rightarrow \langle Agt\rangle(\mathsf{G}\varphi \wedge \mathsf{F}\psi)$$
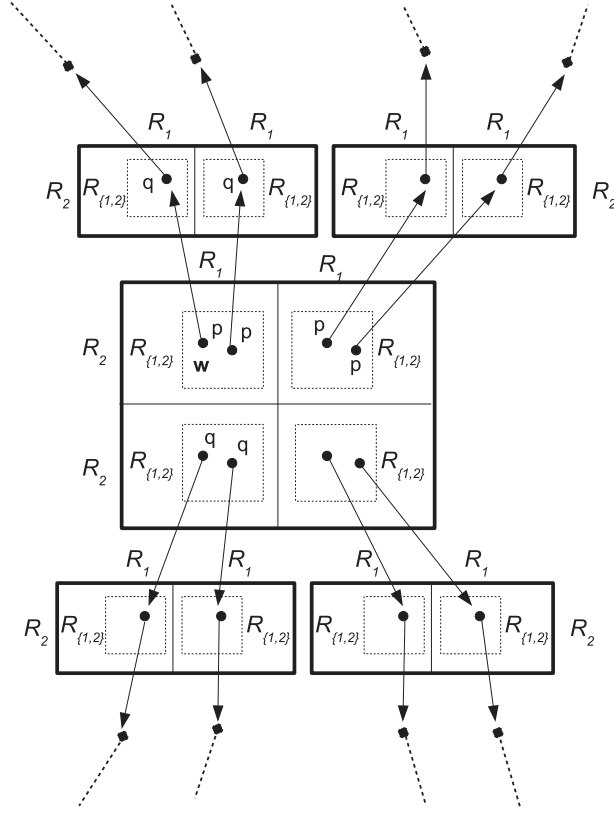
Figure 1. Example of temporal Kripke STIT model.
Note: The world w is the actual world. Choices of agent 1 are represented by columns whereas choices of agent 2 are represented by rows. Choices of the group $\{1, 2\}$ are represented by dotted rectangles. The temporal relation $\mathcal{R}_{\mathsf{G}}$ is represented by the arrows.

Proof. I only prove the first validity, as the second one can be proved in a similar way. I prove its contrapositive, namely I prove that $[Agt]\mathsf{F}\varphi \to \mathsf{F}\Box\mathsf{F}^*\varphi$ is T-STIT valid. Suppose $M, w \models [Agt]\mathsf{F}\varphi$. This means that:

(A) for all $v \in \mathcal{R}_{Agt}(w)$ there is $u \in \mathcal{R}_{\mathsf{G}}(v)$ such that $M, u \models \varphi$.

Let

$$A_w^\varphi = \left\{ u \mid M, u \models \varphi \text{ and } \exists v \in \mathcal{R}_{Agt}(w) \text{ such that } u \in \mathcal{R}_{\mathsf{G}}(v) \right\}$$

and let

$$PA_w^\varphi = \left\{ \bar{u} \in \mathcal{R}_{\mathsf{G}}(w) \mid \exists u \in A_w^\varphi \text{ such that } (\bar{u}, u) \in \mathcal{R}_\Box \right\}.$$

Moreover let

$$NB_w^\varphi = \mathcal{R}_{\mathsf{G}}(w) \setminus \bigcup_{\bar{u} \in PA_w^\varphi} \mathcal{R}_{\mathsf{G}}(\bar{u}).$$

I am going to show that $NB_w^\varphi \neq \emptyset$. Suppose $NB_w^\varphi = \emptyset$. Thus, there exists $\bar{u} \in PA_w^\varphi$ such that $\mathcal{R}_{\mathsf{G}}(w) = \mathcal{R}_{\mathsf{G}}(\bar{u})$. Hence, by Constraint C5, there exists $\bar{u} \in PA_w^\varphi$ such that $w = \bar{u}$.

The latter implies that there exists $\bar{u} \in \mathcal{R}_\mathsf{G}(w)$ such that $w = \bar{u}$, which is in contradiction with the fact that $\mathcal{R}_\mathsf{G}$ is irreflexive.

Take any world $z \in NB_w^\varphi$. Clearly, $z \in \mathcal{R}_\mathsf{G}(w)$. I am going to show that

(B) for all $v \in \mathcal{R}_\square(z)$ either $M, v \models \varphi$ or there is $u \in \mathcal{R}_\mathsf{G}(v)$ such that $M, u \models \varphi$.

Suppose $v \in \mathcal{R}_\square(z)$. From $z \in \mathcal{R}_\mathsf{G}(w)$, by Constraint C6 and the previous observation (A), we have that there are z', v' such that $z' \in \mathcal{R}_{Agt}(w)$, $v \in \mathcal{R}_\mathsf{G}(z')$, $v' \in \mathcal{R}_\mathsf{G}(z')$ and $M, v' \models \varphi$. Moreover, by Constraint C4, $v' \in \mathcal{R}_\mathsf{G}(v)$ or $v \in \mathcal{R}_\mathsf{G}(v')$ or $v = v'$. I am going to show that $v = v'$ or $v' \in \mathcal{R}_\mathsf{G}(v)$ by reductio ad absurdum.

Suppose $v \in \mathcal{R}_\mathsf{G}(v')$. Clearly, $v' \in A_w^\varphi$. From $z' \in \mathcal{R}_{Agt}(w)$, $v \in \mathcal{R}_\square(z)$, $z \in \mathcal{R}_\mathsf{G}(w)$, $v \in \mathcal{R}_\mathsf{G}(z')$, $v' \in \mathcal{R}_\mathsf{G}(z')$ and $v \in \mathcal{R}_\mathsf{G}(v')$, by Proposition 4, it follows that there exists $\bar{v}' \in \mathcal{R}_\mathsf{G}(w)$ such that $v' \in \mathcal{R}_{Agt}(\bar{v}')$ and $z \in \mathcal{R}_\mathsf{G}(\bar{v}')$. Since $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$ and $v' \in A_w^\varphi$, the latter implies that there exists $\bar{v}' \in \mathcal{R}_\mathsf{G}(w)$ such that $\bar{v}' \in PA_w^\varphi$, $v' \in \mathcal{R}_{Agt}(\bar{v}')$ and $z \in \mathcal{R}_\mathsf{G}(\bar{v}')$. $\bar{v}' \in PA_w^\varphi$ and $z \in \mathcal{R}_\mathsf{G}(\bar{v}')$ together imply that $z \notin NB_w^\varphi$. But this is in contradiction with the initial hypothesis that $z \in NB_w^\varphi$.

By the previous item (B), we have that $M, z \models \square\mathsf{F}^*\varphi$. Since $z \in \mathcal{R}_\mathsf{G}(w)$, $M, w \models \mathsf{F}\square\mathsf{F}^*\varphi$. ∎

2.3. Axiomatisation

Figure 2 contains a complete axiomatisation with respect to the class of temporal Kripke STIT models.

PC	All tautologies of classical propositional calculus
S5(i)	All S5-principles for the operators $[i]$
S5($\square$)	All S5-principles for the operator $\square$
S5(Agt)	All S5-principles for the operator $[Agt]$
KD4(G)	All KD4-principles for the operator G
K(H)	All K-principles for the operator H
($\square \to i$)	$\square\varphi \to [i]\varphi$
($i \to Agt$)	$([1]\varphi_1 \wedge \ldots \wedge [n]\varphi_n) \to [Agt](\varphi_1 \wedge \ldots \wedge \varphi_n)$
(AIA)	$(\Diamond[1]\varphi_1 \wedge \ldots \wedge \Diamond[n]\varphi_n) \to \Diamond([1]\varphi_1 \wedge \ldots \wedge [n]\varphi_n)$
(Conv$_{\mathsf{G,H}}$)	$\varphi \to \mathsf{GP}\varphi$
(Conv$_{\mathsf{H,G}}$)	$\varphi \to \mathsf{HF}\varphi$
(Connected$_\mathsf{G}$)	$\mathsf{PF}\varphi \to (\mathsf{P}\varphi \vee \varphi \vee \mathsf{F}\varphi)$
(Connected$_\mathsf{H}$)	$\mathsf{FP}\varphi \to (\mathsf{P}\varphi \vee \varphi \vee \mathsf{F}\varphi)$
(NCUH)	$\mathsf{F}\Diamond\varphi \to \langle Agt\rangle\mathsf{F}\varphi$
(MP)	$\dfrac{\varphi,\ \varphi \to \psi}{\psi}$
(IRR)	$\dfrac{(\square\neg p \wedge \square(\mathsf{G}p \wedge \mathsf{H}p)) \to \varphi}{\varphi}$, provided p does not occur in φ

Figure 2. Axiomatisation of T-STIT.

This includes all tautologies of classical propositional calculus (**PC**) as well as modus ponens (**MP**). Moreover, we have all the principles of the normal modal logic S5 for every operator $[i]$, for the operator $[Agt]$ and for the operator $\Box$, all principles of the normal modal logic KD4 for the future tense operator G and all principles of the normal modal logic K for the past tense operator H. That is, we have Axiom K for each operator: $(\blacksquare\varphi \wedge \blacksquare (\varphi \to \psi)) \to \blacksquare\psi$ with $\blacksquare \in \{\Box, \mathsf{G}, \mathsf{H}, [Agt]\} \cup \{[i]|i \in Agt\}$. We have Axiom D for the future tense modality G: $\neg(\mathsf{G}\varphi \wedge \mathsf{G}\neg\varphi)$. We have Axiom 4 for $\Box$, G, $[Agt]$ and for every $[i]$: $\blacksquare\varphi \to \blacksquare\blacksquare\varphi$ with $\blacksquare \in \{\Box, [Agt], \mathsf{G}\} \cup \{[i]|i \in Agt\}$. Furthermore, we have Axiom T for $\Box$, $[Agt]$ and for every $[i]$: $\blacksquare\varphi \to \varphi$ with $\blacksquare \in \{\Box, [Agt]\} \cup \{[i]|i \in Agt\}$. We have Axiom B for $\Box$, $[Agt]$ and for every $[i]$: $\varphi \to \blacksquare\neg\blacksquare\neg\varphi$ with $\blacksquare \in \{\Box, [Agt]\} \cup \{[i]|i \in Agt\}$. Finally we have the rule of necessitation for each modal operator: $\frac{\varphi}{\blacksquare\varphi}$ with $\blacksquare \in \{\Box, [Agt], \mathsf{G}, \mathsf{H}\} \cup \{[i]|i \in Agt\}$. In what follows, I write $\vdash_{\text{T-STIT}} \varphi$ if φ is a T-STIT theorem. Moreover, I say that φ is T-STIT consistent if $\nvdash_{\text{T-STIT}} \neg\varphi$.

$(\Box \to i)$ and (**AIA**) are the two central principles in Xu's (1998) axiomatisation of the Chellas's STIT operators $[i]$. According to Axiom $(\Box \to i)$, 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.[3] According to Axiom $(i \to Agt)$, all agents bring about together what each of them brings about individually.

We have principles for the tense operators and for the relationship between time and action. (**Connected$_\mathsf{G}$**) and (**Connected$_\mathsf{H}$**) are the basic axioms for the linearity of the future and the linearity of the past (Goldblatt, 1992). (**Conv$_{\mathsf{G,H}}$**) and (**Conv$_{\mathsf{H,G}}$**) are the basic interaction axioms between future and past of minimal tense logic according to which 'what is, will always have been' and 'what is, has always been going to be'.

Axiom (**NCUH**) establishes a fundamental relationship between action and time and corresponds to the semantic constraint of 'no choice between undivided histories' over temporal Kripke STIT models (Constraint C6): if in some future world φ will be possible then the actual collective choice of all agents will possibly result in a state in which φ is true.

(**IRR**) is a variant of the well-known Gabbay's irreflexivity rule that has been widely used in the past for proving completeness results for different kinds of temporal logic in which time is supposed to be irreflexive (see, e.g., Gabbay, Hodkinson, & Reynolds, 1994; Reynolds, 2003; von Kutschera, 1997; Zanardo, 1996). The idea is that the special kind of irreflexivity for the relation $\mathcal{R}_\mathsf{G}$ expressed by the Constraint C7 in Definition 1, although not definable in terms of an axiom, can be characterised in an alternative sense by means of the rule (**IRR**). This rule is perhaps more comprehensible if we consider its contrapositive: if p does not occur in φ and φ is T-STIT consistent, then $\Box\neg p \wedge \Box(\mathsf{G}p \wedge \mathsf{H}p) \wedge \varphi$ is T-STIT consistent.

Theorem 7. *The set of T-STIT validities is completely axiomatised by the principles given in Figure 2.*

2.4. Proof of Theorem 7

It is a routine task to show that the axioms given in Figure 2 are valid with respect to the class of temporal Kripke T-STIT models and that the rules of inference preserve validity. Thus, if φ is a T-STIT theorem then φ is T-STIT valid.

I am going to prove that if φ is T-STIT consistent then φ is T-STIT satisfiable. The proof is divided into two steps.

First of all, I introduce the class of *superadditive* temporal Kripke STIT models. While in the temporal Kripke STIT models the collective choice of the grand coalition is *equal*

to the pointwise intersection of the individual choices, in superadditive temporal Kripke STIT models the collective choice of the grand coalition is merely *included* in the pointwise intersection of the individual choices. I prove that the logic T-STIT does not distinguish the semantics in terms of temporal Kripke STIT models from the more 'liberal' semantics in terms of superadditive temporal Kripke STIT models. That is to say, the set of validities with respect to the class of temporal Kripke STIT models is equal to the set of validities with respect to the class of superadditive temporal Kripke STIT models (Lemma 9).

Secondly, I prove that the set of validities in the class of superadditive temporal Kripke STIT models is completely axiomatised by the principles given in Figure 2 (Lemma 10).

Theorem 7 directly follows from Lemma 10 and Lemma 9.

Let me define the class of *superadditive* temporal Kripke STIT models.

Definition 8 (Superadditive temporal Kripke STIT model). *The class of superadditive temporal Kripke STIT models includes all tuples $M = (W, \mathcal{R}_\square, \{\mathcal{R}_i | i \in Agt\}, \mathcal{R}_{Agt}, \mathcal{R}_\mathsf{G}, \mathcal{R}_\mathsf{H}, \mathcal{V})$ where:*

- *W is a nonempty set of possible worlds;*
- *$\mathcal{R}_\square$, every $\mathcal{R}_i$ and $\mathcal{R}_{Agt}$ are equivalence relations between worlds in W such that:*
  - **(C1)** *$\mathcal{R}_i \subseteq \mathcal{R}_\square$;*
  - **(C2)** *for all $u_1, \ldots, u_n \in W$: if $(u_i, u_j) \in \mathcal{R}_\square$ for all $i, j \in \{1, \ldots, n\}$ then $\bigcap_{1 \leq i \leq n} \mathcal{R}_i(u_i) \neq \emptyset$;*
  - **(C3*)** *for all $w \in W$: $\mathcal{R}_{Agt}(w) \subseteq \bigcap_{i \in Agt} \mathcal{R}_i(w)$;*
- *$\mathcal{R}_\mathsf{G}$ and $\mathcal{R}_\mathsf{H}$ are binary relations between worlds in W such that $\mathcal{R}_\mathsf{G}$ is serial and transitive, $\mathcal{R}_\mathsf{H}$ is the inverse relation of $\mathcal{R}_\mathsf{G}$, and:*
  - **(C4)** *for all $w, v, u \in W$: if $v, u \in \mathcal{R}_\mathsf{G}(w)$ then $u \in \mathcal{R}_\mathsf{G}(v)$ or $v \in \mathcal{R}_\mathsf{G}(u)$ or $u = v$;*
  - **(C5)** *for all $w, v, u \in W$: if $v, u \in \mathcal{R}_\mathsf{H}(w)$ then $u \in \mathcal{R}_\mathsf{H}(v)$ or $v \in \mathcal{R}_\mathsf{H}(u)$ or $u = v$;*
  - **(C6)** *$\mathcal{R}_\mathsf{G} \circ \mathcal{R}_\square \subseteq \mathcal{R}_{Agt} \circ \mathcal{R}_\mathsf{G}$;*
  - **(C7)** *for all $w \in W$: if $v \in \mathcal{R}_\square(w)$ then $v \notin \mathcal{R}_\mathsf{G}(w)$;*
- *$\mathcal{V} : Atm \longrightarrow 2^W$ is a valuation function for atomic formulas.*

The only difference between the class of temporal Kripke STIT models and superadditive temporal Kripke STIT models is in the Constraint C3*.

Lemma 9. *Let φ be a formula in $\mathcal{L}_{\textit{T-STIT}}(Atm, Agt)$. Then, φ is satisfiable in the class of temporal Kripke STIT models if and only if it is satisfiable in the class of superadditive temporal Kripke STIT models.*

Proof. ($\Rightarrow$) The left-to-right direction of the equivalence is obvious as the class of temporal Kripke STIT models is included in the class of superadditive temporal Kripke STIT models.

($\Leftarrow$) As to the right-to-left direction, I am going to show how to transform a superadditive temporal Kripke STIT model into a temporal Kripke STIT model without affecting the satisfiability of a formula. The technique used here is inspired by Vakarelov (1992).

Consider a superadditive temporal Kripke STIT model $M = (W, \mathcal{R}_\square, \{\mathcal{R}_i | i \in Agt\}, \mathcal{R}_{Agt}, \mathcal{R}_\mathsf{G}, \mathcal{R}_\mathsf{H}, \mathcal{V})$ and a world $\underline{w}$ in M such that $M, \underline{w} \models \varphi$. I am going to define a temporal Kripke STIT model $M' = (W', \mathcal{R}'_\square, \{\mathcal{R}'_i | i \in Agt\}, \mathcal{R}'_{Agt}, \mathcal{R}'_\mathsf{G}, \mathcal{R}'_\mathsf{H}, \mathcal{V}')$ that satisfies φ.

For every $i \in Agt$, let

$$\Delta_i = \{\mathcal{R}_i(w) | w \in W\}$$

be the partition of W induced by the equivalence class $\mathcal{R}_i$. Elements of Δ_i are called i's choices. Moreover, let

$$\Delta = \{\delta \in \prod_{i \in Agt} \Delta_i \mid \bigcap_{i \in Agt} \delta_i \neq \emptyset\}$$

be the set of collective choices. Elements of Δ are denoted by $\delta, \delta', \ldots$. For every $\delta \in \Delta$, δ_i denotes the element in the vector δ corresponding to the agent i. Furthermore, for notational convenience I write δ^w to denote the collective choice in Δ that includes the world w. That is, δ^w is the collective choice in Δ such that $w \in \bigcap_{i \in Agt} \delta_i^w$.

For every $\delta \in \Delta$, let

$$\Gamma_\delta = \{\mathcal{R}_{Agt}(w) \mid w \in \bigcap_{i \in Agt} \delta_i\}$$

be the set of $\mathcal{R}_{Agt}$-equivalence classes that are included in $\bigcap_{i \in Agt} \delta_i$.

For every $\delta \in \Delta$, let

$$k_\delta : [1, \ldots, card(\Gamma_\delta)] \longrightarrow \Gamma_\delta$$

be a *bijection* associating every integer between 1 and $card(\Gamma_\delta)$ to a unique element of Γ_δ. In the sequel I write Γ_δ^n to indicate the element $k_\delta(n)$ for every $1 \leq n \leq card(\Gamma_\delta)$.

For every $w \in W$ let

$$TC_w = \{\delta \in \Delta \mid \exists v \in \bigcap_{i \in Agt} \delta_i \text{ such that } v \in \mathcal{T}(w)\}$$

be the set of collective choices that are temporally related with w, with $\mathcal{T}(w) = \mathcal{R}_{\mathsf{H}}(w) \cup \{w\} \cup \mathcal{R}_{\mathsf{G}}(w)$.

Moreover, let

$$PC_w = \{\delta \in \Delta \mid \exists v \in \bigcap_{i \in Agt} \delta_i \text{ such that } v \in \mathcal{R}_{\mathsf{H}}(w)\}$$

be the set of collective choices that are in the past of w.

Finally, let

$$FC_w = \{\delta \in \Delta \mid \exists v \in \bigcap_{i \in Agt} \delta_i \text{ such that } v \in \mathcal{R}_{\mathsf{G}}(w)\}$$

be the set of collective choices that are in the future of w. Note that:

(A1) if $v \in \mathcal{T}(w)$ then $TC_w = TC_v$;

(A2) if $v \in \mathcal{R}_\square(w)$ then $PC_w = PC_v$ (because of Proposition 4. in Section 2.2);

(A3) if $v \in \mathcal{R}_{Agt}(w)$ then $\delta^w = \delta^v$ and $PC_w = PC_v$ (because of (A2)).

Moreover, let Λ_w be the set of all total functions $f : TC_w \longrightarrow \mathbb{Z}^n$ that satisfy the following constraint:

- for all $\delta \in TC_w$ and for all $\overrightarrow{x} \in \mathbb{Z}^n$, if $f(\delta) = \overrightarrow{x}$ then there exists $v \in \mathcal{T}(w)$ such that $v \in \Gamma_\delta^{\Sigma_{x_i \in \{x_1, \ldots, x_n\}} x_i}$,

where $\mathbb{Z}$ is the set of integers and $\overrightarrow{x} = \langle x_1, \ldots, x_n \rangle$. I write $f_i(\delta)$ to denote the i-element in the vector $f(\delta)$, with $1 \leq i \leq n$.

Because of the Constraint C7, we have that for all $w \in W$ and for all $\overrightarrow{x} \in \mathbb{Z}^n$:

(B) if $v \in \mathcal{T}(w)$ and $v \in \Gamma_{\delta^w}^{\Sigma_{x_i \in \{x_1, \ldots, x_n\}} x_i}$ then $w = v$.

Furthermore, because of Proposition 4 in Section 2.2, we have that for all $w, v \in W$ and for all $\overrightarrow{x} \in \mathbb{Z}^n$:

(C) if $v \in \mathcal{R}_\square(w)$ then there exists $u_1 \in \mathcal{R}_H(w)$ such that $u_1 \in \Gamma_\delta^{\Sigma_{x_i \in \{x_1,\dots,x_n\}} x_i}$ if and only if there exists $u_2 \in \mathcal{R}_H(v)$ such that $u_2 \in \Gamma_\delta^{\Sigma_{x_i \in \{x_1,\dots,x_n\}} x_i}$.

From the previous item (C) it follows that for all $w, v \in W$:

(D) if $f \in \Lambda_w$ and $v \in \mathcal{R}_\square(w)$ then there exists $f' \in \Lambda_v$ such that, for all $\delta \in PC_w$, $f'(\delta) = f(\delta)$.

Moreover, from the previous item (A1), for all $w, v \in W$ we have that:

(E) if $f \in \Lambda_w$ and $v \in \mathcal{T}(w)$ then there exists $f' \in \Lambda_v$ such that, for all $\delta \in TC_w$, $f'(\delta) = f(\delta)$.

I am now able to define the model $M' = (W', \mathcal{R}'_\square, \{\mathcal{R}'_i | i \in Agt\}, \mathcal{R}'_{Agt}, \mathcal{R}'_G, \mathcal{R}'_H, \mathcal{V}')$:

- $W' = \{w_f | w \in W \text{ and } f \in \Lambda_w\}$;
- for all $w_f, v_{f'} \in W'$,
  $(w_f, v_{f'}) \in \mathcal{R}'_\square$ iff $(w, v) \in \mathcal{R}_\square$ and $f(\delta) = f'(\delta)$ for all $\delta \in PC_w$;
- for all $i \in Agt$ and for all $w_f, v_{f'} \in W'$,
  $(w_f, v_{f'}) \in \mathcal{R}'_i$ iff $\delta_i^w = \delta_i^v$, $f_i(\delta^w) = f'_i(\delta^v)$ and $f(\delta) = f'(\delta)$ for all $\delta \in PC_w$;
- for all $w_f, v_{f'} \in W'$,
  $(w_f, v_{f'}) \in \mathcal{R}'_{Agt}$ iff $\delta^w = \delta^v$, $f(\delta^w) = f'(\delta^v)$ and $f(\delta) = f'(\delta)$ for all $\delta \in PC_w$;
- for all $w_f, v_{f'} \in W'$,
  $(w_f, v_{f'}) \in \mathcal{R}'_G$ iff $(w, v) \in \mathcal{R}_G$ and $f(\delta) = f'(\delta)$ for all $\delta \in TC_w$;
- for all $w_f, v_{f'} \in W'$,
  $(w_f, v_{f'}) \in \mathcal{R}'_H$ iff $(v_{f'}, w_f) \in \mathcal{R}'_G$;
- for all $p \in Atm$, $\mathcal{V}'(p) = \{w_f \in W' | w \in \mathcal{V}(p)\}$.

It is a routine task to check that the mapping $f : w_f \mapsto w$ defines a *bounded morphism* from M' to M (Blackburn, De Rijke, & Venema, 2001, Definition 2.12). Indeed, it follows from the definitions of $\mathcal{R}'_\square, \mathcal{R}'_i, \mathcal{R}'_{Agt}, \mathcal{R}'_G$ and $\mathcal{R}'_H$ that for all $w_f, v_{f'} \in W'$:

- $(w_f, v_{f'}) \in \mathcal{R}'_i$ implies $(w, v) \in \mathcal{R}_i$;
- $(w_f, v_{f'}) \in \mathcal{R}'_{Agt}$ implies $(w, v) \in \mathcal{R}_{Agt}$;
- $(w_f, v_{f'}) \in \mathcal{R}'_\square$ implies $(w, v) \in \mathcal{R}_\square$;
- $(w_f, v_{f'}) \in \mathcal{R}'_G$ implies $(w, v) \in \mathcal{R}_G$;
- $(w_f, v_{f'}) \in \mathcal{R}'_H$ implies $(w, v) \in \mathcal{R}_H$.

For instance, suppose that $(w_f, v_{f'}) \in \mathcal{R}'_i$. By definition of $\mathcal{R}'_i$ the latter implies that $\delta_i^w = \delta_i^v$ and $w \in \delta_i^w$ and $v \in \delta_i^v$. The latter implies that $(w, v) \in \mathcal{R}_i$.

Now, suppose that $(w_f, v_{f'}) \in \mathcal{R}'_{Agt}$. This implies that $\delta^w = \delta^v$ and $f(\delta^w) = f'(\delta^v)$. By the previous observation (B), it follows that $w, v \in \Gamma_{\delta^w}^{f_1(\delta^w) + \dots + f_n(\delta^w)}$. Thus, $(w, v) \in \mathcal{R}_{Agt}$.

The other way around we have that for all $w_f \in W'$:

- if $(f(w_f), v) \in \mathcal{R}_\square$ then there is $v_{f'}$ such that $(w_f, v_{f'}) \in \mathcal{R}'_\square$;
- if $(f(w_f), v) \in \mathcal{R}_i$ then there is $v_{f'}$ such that $(w_f, v_{f'}) \in \mathcal{R}'_i$;
- if $(f(w_f), v) \in \mathcal{R}_{Agt}$ then there is $v_{f'}$ such that $(w_f, v_{f'}) \in \mathcal{R}'_{Agt}$;
- if $(f(w_f), v) \in \mathcal{R}_G$ then there is $v_{f'}$ such that $(w_f, v_{f'}) \in \mathcal{R}'_G$;
- if $(f(w_f), v) \in \mathcal{R}_H$ then there is $v_{f'}$ such that $(w_f, v_{f'}) \in \mathcal{R}'_H$.

Let me prove the previous items. The first item is trivial and I do not prove it here.

Suppose that $w_f \in W'$ and $(f(w_f), v) \in \mathcal{R}_i$. This implies that $(w, v) \in \mathcal{R}_i$. Therefore, $\mathcal{R}_i(w) = \mathcal{R}_i(v)$ as $\mathcal{R}_i$ is an equivalence relation. Hence, $\delta_i^v = \delta_i^w$. It follows that $(w_f, v_{f'}) \in \mathcal{R}_i'$ where the function f' is defined as follows: (i) $f_i'(\delta^v) = f_i(\delta^w)$ and all $f_j'(\delta^v)$ with $j \neq i$ are such that $v \in \Gamma_{\delta^v}^{f_1'(\delta^v)+...+f_n'(\delta^v)}$; (ii) for all $\delta \in PC_w$, $f'(\delta) = f(\delta)$; (iii) for all $\delta \in FC_v$, $f'(\delta)$ is such that there exists $u \in \mathcal{R}_G(v)$ with $u \in \Gamma_\delta^{f_1'(\delta)+...+f_n'(\delta)}$. This function f' is guaranteed to exist because of the observation (D) above and the fact that $\mathcal{R}_i \subseteq \mathcal{R}_\square$.

Suppose that $w_f \in W'$ and $(f(w_f), v) \in \mathcal{R}_{Agt}$. This implies that $(w, v) \in \mathcal{R}_{Agt}$. Therefore, $\mathcal{R}_{Agt}(w) = \mathcal{R}_{Agt}(v)$ as $\mathcal{R}_{Agt}$ is an equivalence relation. Hence, $\delta^v = \delta^w$. It follows that $(w_f, v_{f'}) \in \mathcal{R}_{Agt}'$ where the function f' is defined as follows: (i) $f'(\delta^v) = f(\delta^w)$; (ii) for all $\delta \in PC_w$, $f'(\delta) = f(\delta)$; (iii) for all $\delta \in FC_v$, $f'(\delta)$ is such that there exists $u \in \mathcal{R}_G(v)$ with $u \in \Gamma_\delta^{f_1'(\delta)+...+f_n'(\delta)}$. This function f' is guaranteed to exist because of the observation (D) above and the fact that $\mathcal{R}_{Agt} \subseteq \mathcal{R}_\square$.

Suppose that $w_f \in W'$ and $(f(w_f), v) \in \mathcal{R}_G$. This implies that $(w, v) \in \mathcal{R}_G$. It follows that $(w_f, v_{f'}) \in \mathcal{R}_G'$ where the function f' is such that, for all $\delta \in TC_w$, $f'(\delta) = f(\delta)$. This function f' is guaranteed to exist because of the observation (E) above. Thus, $(w_f, v_{f'}) \in \mathcal{R}_G'$.

Finally, suppose that $w_f \in W'$ and $(f(w_f), v) \in \mathcal{R}_H$. This implies that $(w, v) \in \mathcal{R}_H$ which is equivalent to $(v, w) \in \mathcal{R}_G$. It follows that $(v_{f'}, w_f) \in \mathcal{R}_G'$ where the function f' is such that, for all $\delta \in TC_v$, $f(\delta) = f'(\delta)$. This function f' is guaranteed to exist because of the observation (E) above. Thus, $(w_f, v_{f'}) \in \mathcal{R}_H'$.

As f is a bounded morphism it holds that $M, \underline{w} \models \varphi$ if and only if $M', \underline{w}_f \models \varphi$. Thus, $M', \underline{w}_f \models \varphi$ for all $\underline{w}_f \in W'$.

In order to terminate the proof we also need to be sure that M' is a temporal Kripke STIT model in the sense of Definition 1. It is a routine task to show that $\mathcal{R}_i'$, $\mathcal{R}_{Agt}'$ and $\mathcal{R}_\square'$ are equivalence relations, that the model transformation preserves transitivity and seriality for the relation $\mathcal{R}_G'$, and that the model M' satisfies the Constraints C1, C2, C4, C5 and C7. Let me prove that it also satisfies Constraints C3 and C6.

$v_{f'} \in \bigcap_{i \in Agt} \mathcal{R}_i'(w_f)$ if and only if (i) $\delta_i^w = \delta_i^v$ and $f_i(\delta^w) = f_i'(\delta^v)$ for all $i \in Agt$; and (ii) $f(\delta) = f'(\delta)$ for all $\delta \in PC_w$. The latter is equivalent to $\delta^w = \delta^v$, $f(\delta^w) = f'(\delta^v)$ and $f(\delta) = f'(\delta)$ for all $\delta \in PC_w$, which in turn is equivalent to $v_{f'} \in \mathcal{R}_{Agt}'(w_f)$. This proves that the model M' satisfies the Constraint C3.

Now suppose that $v_{f'} \in \mathcal{R}_G'(w_f)$ and $u_{f''} \in \mathcal{R}_\square'(v_{f'})$. It follows that: (i) $f(\delta) = f'(\delta)$ for all $\delta \in TC_w$; and (ii) $f'(\delta) = f''(\delta)$ for all $\delta \in PC_v$. Moreover, since the model M satisfies the Constraint C6, there exists $z \in W$ such that $z \in \mathcal{R}_{Agt}(w)$ and $u \in \mathcal{R}_G(z)$. Let $f''' \in \Lambda_z$ be the function that for all $\delta \in TC_z$, $f''(\delta) = f'''(\delta)$. It follows that $u_{f''} \in \mathcal{R}_G'(z_{f'''})$. I am going to prove that $z_{f'''} \in \mathcal{R}_{Agt}'(w_f)$. From the previous conditions (i) and (ii) it follows that $f(\delta) = f''(\delta)$ for all $\delta \in TC_w \cap PC_v$. Since $v \in \mathcal{R}_G(w)$, $TC_w \cap PC_v = PC_v$. Hence, $f(\delta) = f''(\delta)$ for all $\delta \in PC_v$. Furthermore, $f(\delta) = f''(\delta)$ for all $\delta \in PC_w \cup \{\delta^w\}$, because $v \in \mathcal{R}_G(w)$. From the latter and the fact that $f''(\delta^w) = f'''(\delta^z)$ for all $\delta \in TC_z$, it follows that $f(\delta^w) = f'''(\delta^z)$ for all $\delta \in TC_z \cap (PC_w \cup \{\delta^w\})$. Since $z \in \mathcal{R}_{Agt}(w)$, by the previous observation (A3), it follows that $\delta^w = \delta^z$ and $f(\delta^w) = f'''(\delta^z)$ for all $\delta \in PC_w \cup \{\delta^w\}$. Hence, $f(\delta^w) = f'''(\delta^z)$ and $f(\delta) = f'''(\delta)$ for all $\delta \in PC_w$. Thus, $z_{f'''} \in \mathcal{R}_G'(w_{f'})$. This proves that the model M' satisfies the Constraint C6. ∎

The following lemma provides an axiomatisation result for the class of superadditive temporal Kripke STIT models.

Lemma 10. *The set of T-STIT formulas that are valid in the class of superadditive temporal Kripke STIT models is completely axiomatised by the principles given in Figure 2.*

Proof. In order to prove Lemma 10, I use a technique similar to the one used by Gabbay, Hodkinson and Reynolds (1994) for proving completeness of linear temporal logic and of logic of historical necessity by means of a slightly different variant of the *irreflexivity rule* **(IRR)**.

Let me start with the following standard definition of canonical model for T-STIT.

Definition 11 (Canonical model for T-STIT). *The canonical model M^c for T-STIT is the tuple $M^c = (W^c, \mathcal{R}^c_\Box, \{\mathcal{R}^c_i | i \in Agt\}, \mathcal{R}^c_{Agt}, \mathcal{R}^c_\mathsf{G}, \mathcal{R}^c_\mathsf{H}, \mathcal{V}^c)$ where:*

- *W^c is the set of all maximal consistent sets of T-STIT formulas (MCSs);*
- *$\mathcal{R}^c_\Box, \mathcal{R}^c_i, \mathcal{R}^c_{Agt}, \mathcal{R}^c_\mathsf{G}$ and $\mathcal{R}^c_\mathsf{H}$ are respectively the canonical relations for $\Box$, $[i]$, $[Agt]$, G and H, that is:*
    - *for all $\Gamma, \Delta \in W^c$, $(\Gamma, \Delta) \in \mathcal{R}^c_\Box$ iff for all formulas ψ, $\psi \in \Delta$ implies $\Diamond\psi \in \Gamma$;*
    - *for all $\Gamma, \Delta \in W^c$ and for all $i \in Agt$, $(\Gamma, \Delta) \in \mathcal{R}^c_i$ iff for all formulas ψ, $\psi \in \Delta$ implies $\langle i\rangle\psi \in \Gamma$;*
    - *for all $\Gamma, \Delta \in W^c$, $(\Gamma, \Delta) \in \mathcal{R}^c_{Agt}$ iff for all formulas ψ, $\psi \in \Delta$ implies $\langle Agt\rangle\psi \in \Gamma$;*
    - *for all $\Gamma, \Delta \in W^c$, $(\Gamma, \Delta) \in \mathcal{R}^c_\mathsf{G}$ iff for all formulas ψ, $\psi \in \Delta$ implies $\mathsf{F}\psi \in \Gamma$;*
    - *for all $\Gamma, \Delta \in W^c$, $(\Gamma, \Delta) \in \mathcal{R}^c_\mathsf{H}$ iff for all formulas ψ, $\psi \in \Delta$ implies $\mathsf{P}\psi \in \Gamma$;*
- *$\mathcal{V}^c$ is the valuation defined by $\mathcal{V}^c(p) = \{\Gamma \in W^c | p \in \Gamma\}$ for all $p \in Atm$.*

Furthermore, let me introduce the notion of 'diamond'-saturation for maximal consistent sets of T-STIT formulas.

Definition 12 (Diamond saturated set of MCSs). *Given a set X of maximal consistent sets of T-STIT formulas, I say that X is a* diamond saturated *set of MCSs if and only if for each $\Gamma \in X$ the following five conditions are satisfied:*

- *for each formula $\Diamond\varphi \in \Gamma$ there is $\Delta \in X$ such that $(\Gamma, \Delta) \in \mathcal{R}^c_\Box$ and $\varphi \in \Delta$;*
- *for each formula $\langle i\rangle\varphi \in \Gamma$ there is $\Delta \in X$ such that $(\Gamma, \Delta) \in \mathcal{R}^c_i$ and $\varphi \in \Delta$;*
- *for each formula $\langle Agt\rangle\varphi \in \Gamma$ there is $\Delta \in X$ such that $(\Gamma, \Delta) \in \mathcal{R}^c_{Agt}$ and $\varphi \in \Delta$;*
- *for each formula $\mathsf{F}\varphi \in \Gamma$ there is $\Delta \in X$ such that $(\Gamma, \Delta) \in \mathcal{R}^c_\mathsf{G}$ and $\varphi \in \Delta$;*
- *for each formula $\mathsf{P}\varphi \in \Gamma$ there is $\Delta \in X$ such that $(\Gamma, \Delta) \in \mathcal{R}^c_\mathsf{H}$ and $\varphi \in \Delta$.*

The following truth lemma is provable in the standard way by induction on φ (see Blackburn et al., 2001, Lemma 4.70).

Lemma 13 (Truth Lemma). *Let X be a diamond saturated set of MCSs. Then for any $\Gamma \in X$ and for any formula φ, $M^c|_X, \Gamma \models \varphi$ if and only if $\varphi \in \Gamma$, where $M^c|_X$ is the submodel of the canonical model M^c induced by X.*

The next step in the proof consists in defining the notion of IRR theory.

For every atom p, let

$$name(p) \stackrel{\text{def}}{=} \Box\neg p \wedge \Box(\mathsf{G}p \wedge \mathsf{H}p).$$

The formula $name(p)$ acts as a sort of 'name' for a given world that ensures irreflexivity of the canonical relation $\mathcal{R}^c_\mathsf{G}$ for the future tense operator G.

Furthermore, let

$$\Theta = \{ \blacklozenge_1 (\psi_1 \wedge \blacklozenge_2 (\psi_2 \wedge \ldots \wedge \blacklozenge_n \psi_n) \ldots) :$$
$$\blacklozenge_1, \ldots, \blacklozenge_n \in \{ \lozenge, \langle 1 \rangle, \ldots, \langle n \rangle, \langle Agt \rangle, \mathsf{F}, \mathsf{P} \} \text{ and } \psi_1, \ldots, \psi_n \in \mathcal{L}_{\text{T-STIT}}(Atm, Agt) \}.$$

Finally, for every $\varphi = \blacklozenge_1 (\psi_1 \wedge \blacklozenge_2 (\psi_2 \wedge \ldots \wedge \blacklozenge_n \psi_n) \ldots) \in \Theta$ and for every atom p, let

$$\varphi(p) \stackrel{\text{def}}{=} \blacklozenge_1 (\psi_1 \wedge \blacklozenge_2 (\psi_2 \wedge \ldots \wedge \blacklozenge_n (\psi_n \wedge name(p))) \ldots).$$

A formula of the form $\varphi(p)$ is also used as a 'name' for a given world. In particular, formulas of the form $\varphi(p)$ provide 'names' for worlds that are reachable by any zig-zagging sequence of $\lozenge$s, $\langle i \rangle$s, $\langle Agt \rangle$s, Fs and Ps.

Definition 14 (IRR theory). *An IRR theory is a maximal consistent set of T-STIT formulas Γ such that:*

- *for some p, $name(p) \in \Gamma$;*
- *if $\varphi \in \Gamma \cap \Theta$ then, for some atom p, $\varphi(p) \in \Gamma$.*

The set of all IRR theories is denoted by **IrrTh***.*

The following lemma highlights that every consistent T-STIT formula is included in at least one IRR theory. The rule of inference **(IRR)** becomes crucial at this point of the proof.

Lemma 15. *Let φ be a consistent T-STIT formula. Then, there exists an IRR theory Γ such that $\varphi \in \Gamma$.*

(Sketch). The lemma is proved analogously to Lemma 6.2.4 in Gabbay et al. (1994, Chapter 6). Since the set *Atm* of propositional atoms is infinite, there exists an infinite number of atomic formulas p not occurring in φ. Therefore, thanks to the **(IRR)** rule, it is straightforward to build step-by-step an IRR theory Γ containing φ. ∎

The following lemma highlights that the set of all IRR theories satisfies the condition of 'diamond'-saturation defined above (Definition 12).

Lemma 16 (Existence Lemma)**.**
Let Γ be an IRR theory. Then:

- *if $\lozenge \varphi \in \Gamma$, then there is an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_{\square}^c$ and $\varphi \in \Delta$;*
- *if $\langle i \rangle \varphi \in \Gamma$, then there is an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_i^c$ and $\varphi \in \Delta$;*
- *if $\langle Agt \rangle \varphi \in \Gamma$, then there is an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_{Agt}^c$ and $\varphi \in \Delta$;*
- *if $\mathsf{F} \varphi \in \Gamma$, then there is an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_{\mathsf{G}}^c$ and $\varphi \in \Delta$;*
- *if $\mathsf{P} \varphi \in \Gamma$, then there is an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_{\mathsf{H}}^c$ and $\varphi \in \Delta$.*

Proof. I only prove the first item. The other items can be proved analogously.
Suppose $\lozenge \varphi \in \Gamma$. Since Γ is an IRR theory, it follows that $\lozenge (\varphi \wedge name(p)) \in \Gamma$ for some atom p. Let $\Delta_0 = \{ \varphi \wedge name(p) \} \cup \{ \psi | \square \psi \in \Gamma \}$. I am going to show that Δ_0 is consistent by reductio ad absurdum. Suppose Δ_0 is not consistent. Then, for some $\psi_1, \ldots, \psi_k \in \{ \psi | \square \psi \in \Gamma \}$, we have:

$$\vdash (\psi_1 \wedge \ldots \wedge \psi_k) \rightarrow \neg (\varphi \wedge name(p)),$$

hence

$$\vdash \square (\psi_1 \wedge \ldots \wedge \psi_k) \rightarrow \square \neg (\varphi \wedge name(p)),$$

and hence $\Box\neg(\varphi \wedge name(p)) \in \Gamma$ (because $\Box(\psi_1 \wedge \ldots \wedge \psi_k) \in \Gamma$). But this is in contradiction with $\Diamond(\varphi \wedge name(p)) \in \Gamma$.

Now, I turn to define an increasing sequence of consistent extensions $(\Delta_n)_{n \geq 0}$ such that $\Delta_n \subseteq \Delta_{n+1}$ for all n.

Assume $\Delta_n \subseteq \Delta_{n+1}$ has been defined and is consistent. Then, either $\Delta_n \cup \{\chi_n\}$ is consistent or $\Delta_n \cup \{\neg\chi_n\}$ is consistent.

Case 1: $\Delta_n \cup \{\neg\chi_n\}$ *consistent.* Take $\Delta_{n+1} = \Delta_n \cup \{\neg\chi_n\}$.

Case 2a: $\Delta_n \cup \{\neg\chi_n\}$ *not consistent,* $\Delta_n \cup \{\chi_n\}$ *consistent and* $\chi_n \notin \Theta$. Take $\Delta_{n+1} = \Delta_n \cup \{\chi_n\}$.

Case 2b: $\Delta_n \cup \{\neg\chi_n\}$ *not consistent,* $\Delta_n \cup \{\chi_n\}$ *consistent and* $\chi_n \in \Theta$. We have that:

$$\Diamond((\varphi \wedge name(p)) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n) \in \Gamma.$$

For, otherwise,

$$\Box(((\varphi \wedge name(p)) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi) \to \neg\chi_n) \in \Gamma$$

and hence, by definition of Δ_0, $((\varphi \wedge name(p)) \wedge \bigwedge_{\psi \in \Delta_n \setminus \Delta_0} \psi) \to \neg\chi_n \in \Delta_n$.
Hence, since $((\varphi \wedge name(p)) \wedge \bigwedge_{\psi \in \Delta_n \setminus \Delta_0} \psi) \in \Delta_n$, $\neg\chi_n \in \Delta_n$.
This is in contradiction with the fact that $\Delta_n \cup \{\chi_n\}$ is consistent.
Thus, since Γ is an IRR theory, for some atom q, we have

$$\Diamond((\varphi \wedge name(p)) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma.$$

It follows that $\Delta_n \cup \{\chi_n\} \cup \{\chi_n(q)\}$ is consistent. For, otherwise, for some $\psi_1, \ldots, \psi_k \in \{\psi \,|\, \Box\psi \in \Gamma\}$, we have:

$$\vdash (\psi_1 \wedge \ldots \wedge \psi_k) \to \neg((\varphi \wedge name(p)) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)),$$

hence

$$\vdash \Box(\psi_1 \wedge \ldots \wedge \psi_k) \to \Box\neg((\varphi \wedge name(p)) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)),$$

and hence $\Box\neg((\varphi \wedge name(p)) \wedge \bigwedge_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma$ (because $\Box(\psi_1 \wedge \ldots \wedge \psi_k) \in \Gamma$).
But this is in contradiction with $\Diamond((\varphi \wedge name(p)) \wedge \bigwedge_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma$.
Let $\Delta_{n+1} = \Delta_n \cup \{\chi_n\} \cup \{\chi_n(q)\}$.
Thus, the sequence $(\Delta_n)_{n \geq 0}$ is defined and, clearly, $\Delta = \bigcup_{n \geq 0} \Delta_n$ is the desired IRR theory such that $(\Gamma, \Delta) \in \mathcal{R}_\Box^c$ and $\varphi \in \Delta$. ∎

The last part of the proof consists in proving that the submodel of the canonical model for T-STIT including all IRR theories, i.e., the model $M^c|_{\textbf{IrrTh}}$, is indeed a superadditive temporal Kripke STIT model. The following two propositions ensure that $M^c|_{\textbf{IrrTh}}$ satisfies the Constraints C2 and C6.

Proposition 17. *Let* $\Gamma_1, \ldots, \Gamma_n$ *be IRR theories such that for all* $1 \leq i, j \leq n$, $(\Gamma_i, \Gamma_j) \in \mathcal{R}_\Box^c$. *Then, there exists an IRR theory* Δ *such that* $(\Gamma_1, \Delta) \in \mathcal{R}_1^c, \ldots, (\Gamma_n, \Delta) \in \mathcal{R}_n^c$.

Proof. In order to simplify the exposition, let us assume that $n = 2$. The general case for any arbitrary n can be proved analogously.

Suppose Γ_1 and Γ_2 are IRR theories such that for all $1 \le i, j \le 2$, $(\Gamma_i, \Gamma_j) \in \mathcal{R}^c_\Box$. I am going to show that there exists an IRR theory Δ such that $(\Gamma_1, \Delta) \in \mathcal{R}^c_1$ and $(\Gamma_2, \Delta) \in \mathcal{R}^c_2$.

Since Γ_1 and Γ_2 are IRR theories, there exists $name(p_1) \in \Gamma_1$ and $name(p_2) \in \Gamma_2$ for some atoms p_1, p_2.

Let $\Delta_0 = \{name(p_1)\} \cup \{name(p_2)\} \cup \{\psi | [1]\psi \in \Gamma_1\} \cup \{\psi | [2]\psi \in \Gamma_2\}$. I am going to show that Δ_0 is consistent by reductio ad absurdum.

Suppose Δ_0 is not consistent. Then, for some $\psi_1, \ldots, \psi_k \in \Delta_0$, we have:

$$\vdash (\psi_1 \wedge \ldots \wedge \psi_k) \to \bot.$$

Let Y denote the set $\{\psi_1, \ldots, \psi_k\}$. Moreover, let $Y_1 = \{\psi \in Y | [1]\psi \in \Gamma_1$ or $\psi = name(p_1)\}$ and $Y_2 = \{\psi \in Y | [2]\psi \in \Gamma_2$ or $\psi = name(p_2)\}$. Clearly, $Y = Y_1 \cup Y_2$ and:

(A) $[1] \bigwedge_{\psi \in Y_1} \psi \in \Gamma_1$.

The previous item (A) follows from these two facts: (i) for every $\psi \in Y_1$ such that $\psi \ne name(p_1)$, $[1]\psi \in \Gamma_1$ (because ψ is of the form $[1]\chi$ and $[1]\chi \to [1][1]\chi$ is a T-STIT theorem); and (ii) $[1]name(p_1) \in \Gamma_1$ (because $name(p_1) \to [1]name(p_1)$ is a T-STIT theorem). Likewise, we have:

(B) $[2] \bigwedge_{\psi \in Y_2} \psi \in \Gamma_2$.

From the definition of the accessibility relation $\mathcal{R}^c_\Box$ and the fact that $(\Gamma_1, \Gamma_2) \in \mathcal{R}^c_\Box$, it follows that $\Diamond[2] \bigwedge_{\psi \in Y_2} \psi \in \Gamma_1$. Moreover, by the T-STIT theorem $\vdash [1]\psi \to \Diamond[1]\psi$, we have $\Diamond[1] \bigwedge_{\psi \in Y_1} \psi \in \Gamma_1$. By the Axiom **(AIA)**, it follows that $\Diamond([1] \bigwedge_{\psi \in Y_1} \psi \wedge [2] \bigwedge_{\psi \in Y_2} \psi) \in \Gamma_1$. Hence, by Lemma 16, there is an IRR theory Δ such that $(\Gamma_1, \Delta) \in \mathcal{R}^c_\Box$ and $([1] \bigwedge_{\psi \in Y_1} \psi \wedge [2] \bigwedge_{\psi \in Y_2} \psi) \in \Delta$. Hence, by Axiom T for $[i]$, there is an IRR theory Δ such that $(\Gamma_1, \Delta) \in \mathcal{R}^c_\Box$ and $(\bigwedge_{\psi \in Y_1} \psi \wedge \bigwedge_{\psi \in Y_2} \psi) \in \Delta$. The latter implies that $\bigwedge_{\psi \in Y} \psi$ is consistent. Hence, $\not\vdash (\psi_1 \wedge \ldots \wedge \psi_k) \to \bot$.

The rest of the proof proceeds in the same way as the proof of Lemma 16. Specifically, I turn to define an increasing sequence of consistent extensions $(\Delta_n)_{n \ge 0}$ such that $\Delta_n \subseteq \Delta_{n+1}$ for all n.

Assume $\Delta_n \subseteq \Delta_{n+1}$ has been defined and is consistent. Then, either $\Delta_n \cup \{\chi_n\}$ is consistent or $\Delta_n \cup \{\neg\chi_n\}$ is consistent.

Case 1: $\Delta_n \cup \{\neg\chi_n\}$ *consistent.* Take $\Delta_{n+1} = \Delta_n \cup \{\neg\chi_n\}$.

Case 2a: $\Delta_n \cup \{\neg\chi_n\}$ *not consistent,* $\Delta_n \cup \{\chi_n\}$ *consistent and* $\chi_n \notin \Theta$. Take $\Delta_{n+1} = \Delta_n \cup \{\chi_n\}$.

Case 2b: $\Delta_n \cup \{\neg\chi_n\}$ *not consistent,* $\Delta_n \cup \{\chi_n\}$ *consistent and* $\chi_n \in \Theta$. We have that:

$$\langle 1 \rangle (name(p_1) \wedge \bigwedge_{\psi \in \Delta_n \backslash \Delta_0} \psi \wedge \chi_n) \in \Gamma_1.$$

For, otherwise,

$$[1]((name(p_1) \wedge \bigwedge_{\psi \in \Delta_n \backslash \Delta_0} \psi) \to \neg\chi_n) \in \Gamma_1$$

and hence, by definition of Δ_0, $(name(p_1) \wedge \bigwedge_{\psi \in \Delta_n \backslash \Delta_0} \psi) \to \neg\chi_n \in \Delta_n$. Hence, since $(name(p_1) \wedge \bigwedge_{\psi \in \Delta_n \backslash \Delta_0} \psi) \in \Delta_n$, $\neg\chi_n \in \Delta_n$. This is in contradiction with the fact that $\Delta_n \cup \{\chi_n\}$ is consistent.

Therefore, for some atom q,

$$\langle 1 \rangle (name(p_1) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma_1.$$

It follows that $\Delta_n \cup \{\chi_n\} \cup \{\chi_n(q)\}$ is consistent. For, otherwise, for some $\psi_1, \dots, \psi_k \in \{\psi | [1]\psi \in \Gamma_1\}$, we have:

$$\vdash (\psi_1 \wedge \dots \wedge \psi_k) \rightarrow \neg(name(p_1) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)),$$

hence

$$\vdash [1](\psi_1 \wedge \dots \wedge \psi_k) \rightarrow [1]\neg(name(p) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)),$$

and hence $[1]\neg(name(p_1) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma_1$ (because $[1](\psi_1 \wedge \dots \wedge \psi_k) \in \Gamma_1$). But this is in contradiction with $\langle 1 \rangle (name(p_1) \wedge \bigwedge\nolimits_{\psi \in \Delta_n \setminus \Delta_0} \psi \wedge \chi_n(q)) \in \Gamma_1$.

Let $\Delta_{n+1} = \Delta_n \cup \{\chi_n\} \cup \{\chi_n(q)\}$.

Thus, the sequence $(\Delta_n)_{n \geq 0}$ is defined and, clearly, $\Delta = \bigcup_{n \geq 0} \Delta_n$ is the desired IRR theory such that $(\Gamma_1, \Delta) \in \mathcal{R}_1^c$ and $(\Gamma_2, \Delta) \in \mathcal{R}_2^c$. $\blacksquare$

Proposition 18. *Let Δ, Γ and Γ' be IRR theories such that $(\Delta, \Gamma) \in \mathcal{R}_{\mathsf{G}}^c$ and $(\Gamma, \Gamma') \in \mathcal{R}_{\square}^c$. Then, there exists an IRR theory Δ' such that $(\Delta, \Delta') \in \mathcal{R}_{Agt}^c$ and $(\Delta', \Gamma') \in \mathcal{R}_{\mathsf{G}}^c$.*

Proof. Suppose Δ, Γ and Γ' are IRR theories such that $(\Delta, \Gamma) \in \mathcal{R}_{\mathsf{G}}^c$ and $(\Gamma, \Gamma') \in \mathcal{R}_{\square}^c$. It follows that $name(p) \in \Delta$ for some atom p. Hence, by definition of $\mathcal{R}_{\mathsf{H}}^c$, $\mathsf{P}name(p) \in \Gamma$. By definition of $\mathcal{R}_{\square}^c$, it follows that $\Diamond \mathsf{P}name(p) \in \Gamma'$. We have the following T-STIT theorem:

$$\vdash \Diamond \mathsf{P}\varphi \rightarrow \mathsf{P}\langle Agt \rangle \varphi. \tag{1}$$

The following is the Hilbert-style proof:

1. $\vdash \Diamond \mathsf{H}[Agt]\varphi \rightarrow \mathsf{HF}\Diamond \mathsf{H}[Agt]\varphi$.

    By Axiom($\mathbf{Conv_{H,G}}$).

2. $\vdash \mathsf{HF}\Diamond \mathsf{H}[Agt]\varphi \rightarrow \mathsf{H}\langle Agt \rangle \mathsf{FH}[Agt]\varphi$.

    By Axiom($\mathbf{NCUH}$), Axiom K and necessitation for H.

3. $\vdash \mathsf{H}\langle Agt \rangle \mathsf{FH}[Agt]\varphi \rightarrow \mathsf{H}\langle Agt \rangle [Agt]\varphi$.

    By Axiom($\mathbf{Conv_{H,G}}$), Axiom K and necessitation for H and agt.

4. $\vdash \mathsf{H}\langle Agt \rangle [Agt]\varphi \rightarrow \mathsf{H}\varphi$.

    By T-STIT theorem $\vdash \langle Agt \rangle [Agt]\varphi \rightarrow \varphi$, Axiom K and necessitation for H.

5. $\vdash \mathsf{P}\varphi \rightarrow \square \mathsf{P}\langle Agt \rangle \varphi$.

    From 1–4.

6. $\vdash \Diamond \mathsf{P}\varphi \rightarrow \Diamond \square \mathsf{P}\langle Agt \rangle \varphi$.

    By 5, Axiom K and necessitation for $\square$.

7. $\vdash \Diamond \square \mathsf{P}\langle Agt \rangle \varphi \rightarrow \square \mathsf{P}\langle Agt \rangle \varphi$.

    By T-STIT theorem $\vdash \Diamond \square \varphi \rightarrow \square \varphi$.

8. $\vdash \square \mathsf{P}\langle Agt \rangle \varphi \rightarrow \mathsf{P}\langle Agt \rangle \varphi$.

    By Axiom T for $\square$.

9. $\vdash \Diamond \mathsf{P}\varphi \rightarrow \mathsf{P}\langle Agt \rangle \varphi$.

    From 6–8.

Thus, by the previous T-STIT theorem 1, $P\langle Agt\rangle name(p) \in \Gamma'$. Hence, by Lemma 16, there exists an IRR theory Δ' such that $\langle Agt\rangle name(p) \in \Delta'$ and $(\Delta', \Gamma') \in \mathcal{R}_G^c$. By the T-STIT theorem $\vdash \langle Agt\rangle name(p) \rightarrow name(p)$, the latter implies that there exists an IRR theory Δ' such that $name(p) \in \Delta'$ and $(\Delta', \Gamma') \in \mathcal{R}_G^c$. I am going to show that $(\Delta, \Delta') \in \mathcal{R}_{Agt}^c$.

Assume $[Agt]\psi \in \Delta$. We have the following T-STIT theorem:

$$\vdash (name(p) \wedge [Agt]\psi) \rightarrow G\square H(name(p) \rightarrow \psi). \tag{2}$$

The following is the Hilbert-style proof:

1. $\vdash p \rightarrow \neg name(p)$.
   By definition of $name(p)$ and Axiom T for $\square$.
2. $\vdash name(p) \rightarrow [Agt](Gp \wedge Hp)$.
   By definition of $name(p)$, Axiom $(\square \rightarrow i)$ and Axiom $(i \rightarrow Agt)$.
3. $\vdash [Agt](Gp \wedge Hp) \rightarrow ([Agt]Gp \wedge [Agt]Hp)$.
   By Axiom K for $[Agt]$.
4. $\vdash ([Agt]Gp \wedge [Agt]Hp) \rightarrow ([Agt]G\neg name(p) \wedge [Agt]H\neg name(p))$.
   By 1, Axiom K and necessitation for $[Agt]$, G and H.
5. $\vdash (name(p) \wedge [Agt]\psi) \rightarrow ([Agt]\psi \wedge [Agt]G\neg name(p) \wedge [Agt]H\neg name(p))$.
   From 2–4.
6. $\vdash ([Agt]\psi \wedge [Agt]G\neg name(p) \wedge [Agt]H\neg name(p)) \rightarrow$
   $([Agt](\neg name(p) \vee \psi) \wedge [Agt]G(\neg name(p) \vee \psi) \wedge [Agt]H(\neg name(p) \vee \psi))$.
   By Axiom K and necessitation for $[Agt]$, G and H.
7. $\vdash ([Agt](\neg name(p) \vee \psi) \wedge [Agt]G(\neg name(p) \vee \psi) \wedge [Agt]H(\neg name(p) \vee \psi)) \rightarrow$
   $[Agt]((\neg name(p) \vee \psi) \wedge G(\neg name(p) \vee \psi) \wedge H(\neg name(p) \vee \psi))$.
   By Axiom K for $[Agt]$.
8. $\vdash [Agt]((\neg name(p) \vee \psi) \wedge G(\neg name(p) \vee \psi) \wedge H(\neg name(p) \vee \psi)) \rightarrow$
   $[Agt]GH(\neg name(p) \vee \psi)$.
   By Axiom (**Connected$_H$**), Axiom K and necessitation for $[Agt]$.
9. $\vdash [Agt]GH(\neg name(p) \vee \psi) \rightarrow G\square H(\neg name(p) \vee \psi)$.
   By Axiom (**NCUH**).
10. $\vdash G\square H(\neg name(p) \vee \psi) \rightarrow G\square H(name(p) \rightarrow \psi)$.
11. $\vdash (name(p) \wedge [Agt]\psi) \rightarrow G\square H(name(p) \rightarrow \psi)$.
    From 5–10.

Since $[Agt]\psi \in \Delta$ and $name(p) \in \Delta$, by the previous T-STIT theorem 2, we have $G\square H(name(p) \rightarrow \psi) \in \Delta$. Thus, by definition of $\mathcal{R}_G^c$ and the fact that $(\Delta, \Gamma) \in \mathcal{R}_G^c$, $\square H(name(p) \rightarrow \psi) \in \Gamma$. By definition of $\mathcal{R}_\square^c$ and the fact that $(\Gamma, \Gamma') \in \mathcal{R}_\square^c$, $H(name(p) \rightarrow \psi) \in \Gamma'$. Finally, by definition of $\mathcal{R}_H^c$ and the fact that $(\Gamma', \Delta') \in \mathcal{R}_H^c$, $name(p) \rightarrow \psi \in \Delta'$. Since $name(p) \in \Delta'$, it follows that $\psi \in \Delta'$. ∎

Lemma 19. $M^c|_{\text{IrrTh}}$ is a superadditive temporal Kripke STIT model.

Proof. First of all, it is a routine task to prove that: (i) the canonical relations $\mathcal{R}_\square^c$, $\mathcal{R}_i^c$ and $\mathcal{R}_{Agt}^c$ are equivalence relations; (ii) the canonical relation $\mathcal{R}_G^c$ is transitive; (iii) the canonical relation $\mathcal{R}_H^c$ is the inverse of the canonical relation $\mathcal{R}_G^c$; and (iv) the canonical model M^c satisfies the Constraints C1, C3*, C4 and C5.

Indeed, Axioms T, 4 and B for $\square$, $[i]$ and for $[Agt]$ are canonical for reflexivity, transitivity and symmetry thereby ensuring that the canonical relations $\mathcal{R}_\square^c$, $\mathcal{R}_i^c$ and $\mathcal{R}_{Agt}^c$ are equivalence relations. Axiom 4 for G is canonical for transitivity, thereby ensuring that the canonical relation $\mathcal{R}_G^c$ is transitive. Axiom (**Conv$_{G,H}$**) is canonical for the condition $\mathcal{R}_H \subseteq \mathcal{R}_G^{-1}$, while Axiom (**Conv$_{H,G}$**) is canonical for the condition $\mathcal{R}_G \subseteq \mathcal{R}_H^{-1}$, thereby ensuring that $\mathcal{R}_H^c$ is the inverse of the canonical relation $\mathcal{R}_G^c$.

Finally, Axioms ($\square \rightarrow i$), ($i \rightarrow Agt$), (**Connected$_G$**) and (**Connected$_H$**) are canonical respectively for the Constraints C1, C3*, C4 and C5, thereby ensuring that the canonical model M^c satisfies them.

Since $M^c|_{\mathbf{IrrTh}}$ is a submodel of the canonical model M^c, $M^c|_{\mathbf{IrrTh}}$ inherits from M^c all previous *universal* properties. Specifically, we have that: (i) the relations $\mathcal{R}_\square^c|_{\mathbf{IrrTh}}$, $\mathcal{R}_i^c|_{\mathbf{IrrTh}}$ and $\mathcal{R}_{Agt}^c|_{\mathbf{IrrTh}}$ are equivalence relations; (ii) the relation $\mathcal{R}_G^c|_{\mathbf{IrrTh}}$ is transitive; (iii) the relation $\mathcal{R}_H^c|_{\mathbf{IrrTh}}$ is the inverse of the relation $\mathcal{R}_G^c|_{\mathbf{IrrTh}}$; and (iv) the model $M^c|_{\mathbf{IrrTh}}$ satisfies the Constraints C1, C3*, C4 and C5.

By Propositions 17 and 18, $M^c|_{\mathbf{IrrTh}}$ also satisfies Constraints C2 and C6.

It remains to prove that the relation $\mathcal{R}_G^c|_{\mathbf{IrrTh}}$ is serial and that the model $M^c|_{\mathbf{IrrTh}}$ satisfies the Constraint C7.

Since $G\top$ is a T-STIT theorem, every IRR theory contains it. Thus, by Lemma 16, for every IRR theory Γ there exists an IRR theory Δ such that $(\Gamma, \Delta) \in \mathcal{R}_G^c$. This guarantees that the relation $\mathcal{R}_G^c|_{\mathbf{IrrTh}}$ is serial.

Now, suppose that $(\Gamma, \Delta) \in \mathcal{R}_\square^c|_{\mathbf{IrrTh}}$. Since Δ is an IRR theory, $name(p) \in \Delta$ for some atom p. Hence, $p \in \Delta$. Moreover, $\square G \neg p \in \Delta$. By definition of $\mathcal{R}_\square^c$, it follows that $\Diamond \square G \neg p \in \Gamma$. Thus, by the T-STIT theorem $\vdash \Diamond \square \varphi \rightarrow \square \varphi$, $\square G \neg p \in \Gamma$. By Axiom T for $\square$, it follows that $G \neg p \in \Gamma$. The latter implies that $Fp \notin \Gamma$. Since $p \in \Delta$, this guarantees that $(\Gamma, \Delta) \notin \mathcal{R}_G^c|_{\mathbf{IrrTh}}$. Thus, $M^c|_{\mathbf{IrrTh}}$ satisfies the Constraint C7. ∎

Lemma 10 follows from Lemma 13, Lemma 15, Lemma 16 and Lemma 19. Indeed, suppose that φ is a T-STIT consistent formula. Then, by Lemma 15, there exists an IRR theory Γ in the model $M^c|_{\mathbf{IrrTh}}$ such that $\varphi \in \Gamma$. By Lemma 19, model $M^c|_{\mathbf{IrrTh}}$ is a superadditive temporal Kripke STIT model. Moreover, by Lemma 16, the set of all IRR theories is diamond saturated. Thus, by Lemma 13, $M^c|_{\mathbf{IrrTh}}, \Gamma \models \varphi$. It follows that φ is T-STIT satisfiable. ∎

Theorem 7 follows from Lemma 9 and Lemma 10.

3. Application to normative reasoning

Many normative concepts such as achievement obligation, obligation with deadline and social commitment have an intrinsic agentive and temporal nature – that is to say, they cannot be properly understood without considering their relationships with the concepts of action and time. The aim of this section is to show that the logic T-STIT is expressive enough to capture some of these relationships. I focus on the notion of social commitment and on its relationship with the notion of achievement obligation, postponing the logical analysis of obligations with deadline in STIT to future work.

According to Singh (1999) and Castelfranchi (1995), a social commitment is a kind of normative relationship between a *debtor* and a *creditor*. The contexts in which commitments

are undertaken and established are often institutional contexts. For instance, after signing a contract in the presence of a public notary, a person becomes committed in front of the State to carry out her part of the contract. In this article, I only consider pragmatic commitments and I leave aside propositional commitments (also called dialectical commitments). Pragmatic commitments are about what is to be done whereas propositional commitments are about what is true. Pragmatic commitments concern promises from a debtor to a creditor to perform a given action, while propositional commitments are about positions taken during a dialogue. For example, if i tells to j: 'I will lend you my car for the weekend!' then, he makes a pragmatic commitment to j. On the contrary, if i tells to j: 'Tomorrow, will be sunny. I am sure!' then, he makes a propositional commitment to j.

In order to be able to define social commitment, let us suppose that the set of propositional atoms *Atm* contains special atoms of the form $v_{i,j}$, one for every $i, j \in Agt$ such that $i \neq j$. These are similar to the special atoms for *violation* that are used in the Anderson's reduction of deontic logic to alethic logic (Anderson, 1958; Lindahl, 1994). In the semantics, the special atom $v_{i,j}$ is used to identify those worlds in which agent i does not fulfil her commitments to agent j. The atom $v_{i,j}$ has to be read 'in the actual world, agent i does not fulfil her commitment to agent j'. Special atoms $v_{i,j}$ can be combined with the tense operator G^* in order to identify those histories in which agent i will never fulfil her commitment to agent j. In particular, formula $G^*v_{i,j}$ has to be read 'in the actual history, agent i will never fulfil her commitment to agent j' or also 'in the actual history, agent j is wronged by agent i'.

I say that agent i is committed to agent j to ensure φ (denoted by $\mathbf{C}_{i:j}\varphi$) if and only if: (i) all historic alternatives in which i will never see to it that φ, are histories in which agent i will never fulfil her commitment to agent j; and (ii) i does not see to it that φ. In other words, i will not fulfil her commitment to j unless i will see to it that φ at some point in the future. For every $i, j \in Agt$, I define:

$$\mathbf{C}_{i:j}\varphi \stackrel{\text{def}}{=} \Box(\neg F^*[i]\varphi \rightarrow G^*v_{i,j}) \wedge \neg[i]\varphi.$$

Let me make three observations about the preceding definition of commitment. First of all, it is worth noting that it is related to the notion of achievement obligation (Broersen et al., 2003; Governatori & Rotolo, 2010). The idea is that an agent has an achievement obligation to bring about φ if and only if she has the obligation to bring about φ at some point in the future.[4] Thus, commitment can be seen as a specific kind of achievement obligation. Specifically, it can be conceived as a *directed achievement obligation* from a bearer to a counterparty (for some analysis of the notion of directed obligation in deontic logic, see, e.g., Dignum, 1999; Herrestad & Krogh, 1995; Kanger & Kanger, 1966; Lindahl, 1994; Makinson, 1986).[5] That is to say, agent i is committed to agent j to bring about φ if and only if i has the obligation *towards* j to bring about φ at some point in the future. Secondly, the component $\neg[i]\varphi$ expresses that i is committed to j to ensure φ only if i has not yet fulfilled her commitment to j by bringing about φ. In this sense, the formula $[i]\varphi$ can be conceived as the *discharge* condition for commitment. Suppose agent i is committed to agent j to ensure φ. Then, if i sees to it that φ, then her commitment to j is discharged and is no longer active. Finally, according to the preceding definition, the fact that i brought about φ in the past (before being committed to j) is irrelevant. Indeed, the general intuition is that being committed *now* to do some action means being obliged to do the action either in the present or at some point in the strict future. Therefore, while the fact of performing the action in the present is a sufficient condition for discharging an actual commitment, the fact of having performed the action in the past is not.

The following example illustrates the preceding definition of commitment in a concrete scenario.

Example 20. *Agent 2 is the programme chair of a given conference. Agent 2 asks agent 1, a member of the programme committee, to review some articles submitted for the conference. Agent 1 accepts agent 2's request by sending a confirmation e-mail (we suppose that the communication between 1 and 2 is made through the Easychair system). Consequently, according to the programme committee of the conference, agent 1 is committed to 2 to review the articles: $\mathbf{C}_{1:2}$ review. This means that (i) all historic alternatives in which 1 will never review the articles are histories in which 2 is wronged by 1; and (ii) 1 has not yet fulfilled her commitment to 2 by reviewing the articles:*

$$\Box(\neg\mathsf{F}^*[1]review \to \mathsf{G}^*\mathrm{v}_{1,2}) \wedge \neg[1]review.$$

Now, let us consider some logical properties of the commitment operator $\mathbf{C}_{i:j}$. According to the following T-STIT theorem 3, if i is committed to j to ensure $\varphi \wedge \psi$ then i is committed to j to ensure φ and to ensure ψ. For every $i, j \in Agt$ we have:

$$\vdash_{\text{T-STIT}} \mathbf{C}_{i:j}(\varphi \wedge \psi) \to (\mathbf{C}_{i:j}\varphi \wedge \mathbf{C}_{i:j}\psi). \tag{3}$$

Note that the converse of T-STIT theorem 3 is not valid. Indeed, the fact that i is committed to j to ensure φ and to ensure ψ (i.e., $\mathbf{C}_{i:j}\varphi \wedge \mathbf{C}_{i:j}\psi$) does not necessarily imply that i is committed to j to ensure that φ and ψ are true at the same point in the future (i.e., $\mathbf{C}_{i:j}(\varphi \wedge \psi)$). For example, i may be committed to j to lend her a car and to lend her a motorbike at different points in the future without being committed to lending her a car and a motorbike at the same point in the future.

The following theorem 4 highlights that an agent cannot be committed to bring about tautologies:

$$\vdash_{\text{T-STIT}} \neg\mathbf{C}_{i:j}\top. \tag{4}$$

Indeed, since $[i]\top$ is a T-STIT theorem, agent i's commitment to bring about $\top$ is always discharged.

The following T-STIT theorem 5 is a *weakening* principle for commitment. A similar property of commitment has been isolated by Singh (2008). For every $i, j \in Agt$ we have:

$$\vdash_{\text{T-STIT}} (\mathbf{C}_{i:j}(\varphi \wedge \psi) \wedge [i]\varphi) \to \mathbf{C}_{i:j}\psi. \tag{5}$$

This means that if agent i is committed to agent j to ensure $\varphi \wedge \psi$ and agent i sees to it that φ, then i is committed to j to ensure ψ. So, if an agent is committed to ensuring two states of affairs φ and ψ and her commitment to ensure φ is discharged, then the agent is committed to ensuring ψ.

The following T-STIT theorem 6 clarifies how the previous definition of commitment behaves in the case of Moore-like sentences of the form $\varphi \wedge \neg[i]\varphi$: an agent is committed to ensure that φ is true and that she does not see to it that φ if and only if the agent is committed to do something inconsistent:

$$\vdash_{\text{T-STIT}} \mathbf{C}_{i:j}(\varphi \wedge \neg[i]\varphi) \leftrightarrow \mathbf{C}_{i:j}\bot. \tag{6}$$

The following T-STIT theorem 7 characterises the relationship between commitments about logical equivalent formulas:

$$\vdash_{\text{T-STIT}} \Box\mathsf{G}^*(\varphi \leftrightarrow \psi) \to (\mathbf{C}_{i:j}\varphi \leftrightarrow \mathbf{C}_{i:j}\psi). \tag{7}$$

This means that if φ and ψ are equivalent in all future worlds of every history passing through the current moment, then agent i is committed to ensuring φ if and only if agent i is committed to ensuring ψ. Note that the formula $\Box G^*\varphi$ captures a form of *future necessity* or, even better, *future inevitability*. Therefore, theorem 7 captures the idea that if in the future it is inevitable that φ and ψ have the same truth value, then one cannot be committed to ensuring φ without being committed to ensuring ψ, and one cannot be committed to ensuring ψ without being committed to ensuring φ.

The last T-STIT theorem considered here is the following one about the relationship between time and commitment:

$$\vdash_{\text{T-STIT}} \mathbf{C}_{i:j}\varphi \rightarrow \mathsf{F}(\neg[i]\varphi \rightarrow \mathbf{C}_{i:j}\varphi). \tag{8}$$

Since the Hilbert-style derivation of this T-STIT theorem is not trivial, I give it here:

1. $\vdash \mathbf{C}_{i:j}\varphi \stackrel{\text{def}}{=} \Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi$.

2. $\vdash (\Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi) \rightarrow$
   $([Agt](\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge [Agt]\neg[i]\varphi)$.
   By Axiom $(\Box \rightarrow i)$, Axiom $(i \rightarrow Agt)$ and T-STIT theorem $\neg[i]\varphi \rightarrow [i]\neg[i]\varphi$.

3. $\vdash ([Agt](\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge [Agt]\neg[i]\varphi) \rightarrow$
   $[Agt]((\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi)$.
   By Axiom K for $[Agt]$.

4. $\vdash [Agt]((\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi) \leftrightarrow$
   $[Agt](([i]\varphi \vee \mathsf{F}[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi)$.

5. $\vdash [Agt](([i]\varphi \vee \mathsf{F}[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j}) \wedge \neg[i]\varphi) \rightarrow$
   $[Agt](\mathsf{F}[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j})$.
   By Axiom K for $[Agt]$.

6. $\vdash [Agt](\mathsf{F}[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j}) \rightarrow [Agt](\mathsf{F}[i]\varphi \vee \mathsf{G}\mathrm{v}_{i,j})$.
   By Axiom K for $[Agt]$.

7. $\vdash [Agt](\mathsf{F}[i]\varphi \vee \mathsf{G}\mathrm{v}_{i,j}) \rightarrow \mathsf{F}\Box(\mathsf{F}^*[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j})$.
   By the second T-STIT validity in Proposition 6.

8. $\vdash \mathsf{F}\Box(\mathsf{F}^*[i]\varphi \vee \mathsf{G}^*\mathrm{v}_{i,j}) \rightarrow \mathsf{F}\Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j})$.

9. $\vdash \mathsf{F}\Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}) \rightarrow$
   $\mathsf{F}(([i]\varphi \vee \neg[i]\varphi) \wedge \Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j}))$.
   By T-STIT theorem $\mathsf{F}\varphi \rightarrow \mathsf{F}(\top \wedge \varphi)$.

10. $\vdash \mathsf{F}(([i]\varphi \vee \neg[i]\varphi) \wedge \Box(\neg\mathsf{F}^*[i]\varphi \rightarrow \mathsf{G}^*\mathrm{v}_{i,j})) \rightarrow$
    $\mathsf{F}([i]\varphi \vee \mathbf{C}_{i:j}\varphi)$.

11. $\vdash \mathbf{C}_{i:j}\varphi \rightarrow \mathsf{F}(\neg[i]\varphi \rightarrow \mathbf{C}_{i:j}\varphi)$.
    From 1–10.

According to the previous T-STIT theorem 8, if agent i is committed to agent j to ensure φ then, at some point in the future, either i will see to it that φ or i will remain committed to j to ensure φ. This captures a kind of *persistence* of commitment over time.[6] That is, a commitment will persist as long as the committed agent does not perform the action to

which she is committed. Note that this property of commitment relies on the assumption that there are no 'external' actions or processes that may cancel a pre-existent commitment independently from the fact that the committed agent performs the action for which she is committed. For instance, getting back to the above example, agent 1's commitment to agent 2 to review the articles could be dropped not only because 1 performs the action for which is committed but also because 2 decides to cancel 1's commitment by assigning to a different programme committee member the reviews that were initially assigned to 1.

4. Related work

The logic T-STIT presented in Section 2 differs in several aspects from the system proposed by Wölf (2002). First of all, the language of T-STIT contains the operator [Agt] for the joint action of the grand coalition, while the language of Wölf's logic does not. I have shown that the operator [Agt] is fundamental in order to characterise a basic property relating action and time studied in STIT: the so-called property of *no choice between undivided histories*. Moreover, different to the system of Wölf, T-STIT does not have extra modal operators such as the *difference* operator $\Box^{\neq}$ taken from de Rijke (1992), where $\Box^{\neq}\varphi$ means that 'φ is true in all possible histories that are different from the actual history', and the modal operator $\boxminus$ taken from Maio and Zanardo (1998), where $\boxminus\varphi$ means that 'φ is true in all possible histories that belong to the current instant'. In this sense, the language of T-STIT is more *minimalistic* than the language of Wölf's logic, as it contains only those operators which are strictly necessary for talking about tense and action in an interesting way, while Wölf's logic contains the preceding two extra modal operators $\Box^{\neq}$ and $\boxminus$, which are not necessary for this. Wölf's system and T-STIT are also different at the semantic level. Wölf gives two semantics for his temporal variant of STIT: one based on so-called $T \times W$-based agent frames, which are extensions of $T \times W$-frames (Thomason, 1984), and the other based on so-called tree-based agent-frames, which are similar to the **BT+AC** structures of Belnap et al. (2001). Wölf proves that the two semantics are equivalent. The advantage of the T-STIT semantics given in Section 2.2, compared to Wölf's two semantics, is that the former is closer to the standard semantics of modal logic (Blackburn et al., 2001), as it is based on the standard notion of Kripke model with accessibility relations. A comparison between the semantics of T-STIT in terms of temporal Kripke STIT models and Wölf's two semantics is deferred to future work.

T-STIT also differs from the system proposed by Broersen (Broersen, 2008a,b) under several aspects. First of all, Broersen presents a variant of STIT called XSTIT in which the temporal dimension and the agency dimension are fused to make up a single modal operator. In particular, in Broersen's logic there are primitive operators describing the effects of an agent's action in 'next' states, where 'next' refers to immediate successors of the present state. On the contrary, in T-STIT, the temporal dimension and the agency dimension are kept separate. Secondly, different to T-STIT, Broersen's logic XSTIT does not have a future tense operator. Thirdly, while the notion of group action used in T-STIT corresponds to the notion of group action given by Horty (2001), the notion of group action in Broersen's XSTIT logic does not. According to Horty's definition, the set of outcomes that is forced by the joint action of a coalition is equal to the pointwise intersection of the sets of outcomes that are forced by the individual actions of the agents in the coalition. In Broersen's logic only the left-to-right direction of Horty's (2001) definition holds (i.e., the set of outcomes

that is forced by the individual action of an agent in a coalition is included in the set of outcomes that is forced by the joint action of the coalition).

Another logic that is related to the logic T-STIT presented in Section 2 is Schwarzentruber's (2012) variant of STIT with discrete time extended with the tense operator 'next' of linear temporal logic (LTL). Schwarzentruber provides complexity results for this logic without considering the issue of axiomatisation. However, the expressive power of Schwarzentruber's temporal variant of STIT is too limited for modelling normative concepts such as achievement obligation and commitment. Indeed, the tense operator 'next' is insufficient to express temporal properties about commitments such as the fact that an agent is committed to doing something *in the future*. On the contrary, as I have shown in Section 3, the tense operator G (henceforth) allows us to express such properties.

5. Conclusion

I have presented in this work a temporal variant of STIT which supports reasoning about the temporal properties of normative concepts such as achievement obligation and commitment. A sound and complete axiomatisation for this logic has been given.

Directions of future work are manifold. I defer to future work an extension of the logic T-STIT by the tense operators *until* and *before* of linear temporal logic (LTL; Gabbay et al., 1980). With these operators it will be possible to provide a definition of commitment based on the notion of *deadline* of the following form: agent i is committed to agent j to ensure φ before the deadline ψ if and only if, if i does not see to it that φ *before* ψ becomes true, then j will be wronged by i.

Another issue that I plan to investigate in future research is a comparison between the Kripke-style semantics of T-STIT given in Section 2 and an alternative semantics for T-STIT based on the original notion of the **BT**+**AC** structures of Belnap et al. (2001). Indeed, while the two semantics are clearly equivalent in the case of atemporal individual STIT and atemporal group STIT, one might wonder whether they lead to two different sets of validities in the case of T-STIT and, if so, whether the validities differentiating the two semantics capture interesting and intuitive properties.

Finally, on the technical side, I intend to provide an alternative axiomatisation of T-STIT which does not make use of the Gabbay-style irreflexivity rule **(IRR)**. Indeed, following Zanardo (1996, Theorem 6.12), it seems possible to find such an alternative axiomatisation in which the rule **(IRR)** is replaced with a set of more complex axiom schemas.

Another issue for future work is decidability of the satisfiability problem of T-STIT. It has been proved by Herzig and Schwarzentruber (2008) that Horty's group STIT with group agency operators for all coalitions is undecidable. However, the logic T-STIT only has the group agency operator for the grand coalition. Because of this limitation in the expressive power of T-STIT, I believe that its satisfiability problem is decidable.

Notes

1. The distinction between the 'Ockhamist' view and the 'Peircean' view of branching time was proposed by Prior's (1967) seminal work on the logic of time. According to the 'Peircean' view, the truth of a temporal formula should be evaluated with respect either to some history or all

histories passing through a given moment. The 'Ockhamist' view considers a notion of *actual course of events*. In particular, according to the 'Ockhamist' view, the truth of a temporal formula should be evaluated with respect to a particular *actual* history passing through a given moment.

2. The function f is an order-isomorphism between $\mathcal{R}_\mathsf{H}(w)$ and $\mathcal{R}_\mathsf{H}(v)$ if and only if f is a bijective function $f : \mathcal{R}_\mathsf{H}(w) \longrightarrow \mathcal{R}_\mathsf{H}(v)$ with the property that for all $u, u' \in \mathcal{R}_\mathsf{H}(w)$, $(u, u') \in \mathcal{R}_\mathsf{H}$ if and only if $(f(u), f(u')) \in \mathcal{R}_\mathsf{H}$.

3. Xu (1998) considers a family of axiom schemas $(\mathbf{AIA}_k)$ for independence of agents of the form $(\Diamond[1]\varphi_1 \wedge \ldots \wedge \Diamond[k]\varphi_k) \rightarrow \Diamond([1]\varphi_1 \wedge \ldots \wedge [k]\varphi_k)$ that is parameterised by the integer k. As pointed out by Belnap et al. (2001), $(\mathbf{AIA}_{k+1})$ implies $(\mathbf{AIA}_k)$. Therefore, as *Agt* is finite, in T-STIT the family of axiom schemas can be replaced by the single axiom $(\mathbf{AIA})$.

4. The notion of achievement obligation is traditionally opposed to the concept of maintenance obligation (i.e., the obligation to maintain a given state of affairs φ).

5. Starting from Hohfeld (1917), in legal theory it is assumed that a *right* for j towards i that φ is brought about by i typically correlates with a directed obligation for i towards j to bring about φ (i.e., a *duty* of i towards j to bring about φ).

6. Note that a similar property of persistence characterises the notion of intention (Bratman, 1987) and has been formally characterised in some logical theories of intention (Cohen & Levesque, 1990; Lorini & Herzig, 2006).

References

Anderson, A. (1958). A reduction of deontic logic to alethic modal logic. *Mind*, *22*, 100–103.

Balbiani, P., Herzig, A., & Troquard, N. (2008). Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic*, *37*, 387–406.

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future: Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Bentahar, J., Moulin, B., Ch. Meyer, J.-J., & Chaib-draa, B. (2004). A logical model for commitment and argument network for agent communication. In N. Jennings & M. Tambe (Eds.), *3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004), 19–23 August 2004, New York, NY, USA, proceedings* (Vol. 2, pp. 792–799). Washington, DC: IEEE Computer Society.

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge: Cambridge University Press.

Bratman, M. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Broersen, J. (2008a). A complete STIT logic for knowledge and action, and some of its applications. In M. Baldoni, T. C. Son, M. B. van Riemsdijk, & M. Winikoff (Eds.), *Declarative Agent Languages and Technologies VI, 6th International Workshop, DALT 2008, Estoril, Portugal, May 12, 2008, revised selected and invited papers* (pp. 47–59). Berlin: Springer.

Broersen, J. (2008b). A logical analysis of the interaction between 'obligation-to-do' and 'knowingly doing'. In R. van der Meyden & L. van der Torre (Eds.), *Deontic Logic in Computer Science, 9th International Conference, DEON 2008, Luxembourg, July 15–18, 2008. Proceedings* (pp. 140–154). Berlin: Springer.

Broersen, J., Dastani, M., & van der Torre, L. (2003). BDIO-CTL: Obligations and the specification of agent behavior. In G. Gottlob & T. Walsh (Eds.), *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9–15, 2003* (pp. 1389–1390). San Francisco, CA: Morgan Kaufmann.

Castelfranchi, C. (1995). Commitment: From individual intentions to groups and organizations. In V. R. Lesser & L. Gasser (Eds.), *Proceedings of the First International Conference on Multiagent Systems, June 12–14, 1995, San Francisco, California, USA* (pp. 528–535). Cambridge, MA: MIT Press.

Chellas, B. J. (1992). Time and modality in the logic of agency. *Studia Logica*, *51*, 485–517.

Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, *42*, 213–261.

de Rijke, M. (1992). The modal logic of inequality. *The Journal of Symbolic Logic*, *57*, 566–584.

Desai, N., Narendra, N. C., & Singh, M. P. (2008). Checking correctness of business contracts via commitments. In L. Padgham, D. C. Parkes, J. P. Müller, & S. Parsons (Eds.), *7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), Estoril, Portugal, May 12–16, 2008* (Vol. 2, pp. 787–794). New York, NY: ACM Press.

Dignum, F. (1999). Autonomous agents with norms. *Artificial Intelligence and Law*, *7*, 69–79.

Gabbay, D. M., Hodkinson, I. M., & Reynolds, M. A. (1994). *Temporal logic: Mathematical foundations and computational aspects* (Vol. 1). Oxford: Clarendon Press.

Gabbay, D. M., Pnueli, A., Shelah, S., & Stavi, J. (1980). On the temporal analysis of fairness. In P. W. Abrahams, R. J. Lipton, & S. R. Bourne (Eds.), *Conference Record of the Seventh Annual ACM Symposium on Principles of Programming Languages, Las Vegas, Nevada, USA, January 1980* (pp. 163–173). New York, NY: ACM Press.

Goldblatt, R. (1992). *Logics of time and computation*. (2nd ed.). Stanford, CA: Center for the Study of Language and Information.

Governatori, G., & Rotolo, A. (2010). Norm compliance in business process modeling. In M. Dean, J. Hall, A. Rotolo, & S. Tabet (Eds.), *Semantic Web Rules - International Symposium, RuleML 2010, Washington, DC, USA, October 21–23, 2010. Proceedings* (pp. 194–209). Berlin: Springer.

Herrestad, H., & Krogh, C. (1995). Obligations directed from bearers to counterparties. *Proceedings of the Fifth International Conference on Artificial Intelligence and Law, ICAIL '95, May 21–24, 1995, College Park, Maryland, USA* (pp. 210–218). New York, NY: ACM Press.

Herzig, A., & Schwarzentruber, F. (2008). Properties of logics of individual and group agency. In C. Areces & R. Goldblatt (Eds.), *Advances in Modal Logic* (Vol. 7, pp. 133–149). London: King's College.

Hohfeld, W. (1917). Some fundamental legal conceptions as applied in legal reasoning. *The Yale Law Journal*, *26*, 710–770.

Horty, J. F. (2001). *Agency and deontic logic*. Oxford: Oxford University Press.

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

Kanger, S., & Kanger, H. (1966). Rights and parliamentarism. *Theoria*, *6*, 85–115.

Kooi, B., & Tamminga, A. (2008). Moral conflicts between groups of agents. *Journal of Philosophical Logic*, *37*, 1–21.

Lindahl, L. (1994). Stig Kanger's theory of rights. In D. Prawitz, B. Skyrms, & D. Westeråhl (Eds.), *Logic, methodology and philosophy of science IX*. Philadelphia, PA: Elsevier.

Lorini, E., & Herzig, A. (2006). A logic of intention and attempt. *Synthese*, *163*, 45–77.

Lorini, E., & Schwarzentruber, F. (2011). A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, *175*, 814–847.

Maio, M. C. D., & Zanardo, A. (1998). A Gabbay-rule free axiomatization of T x W validity. *Journal of Philosophical Logic*, *27*, 435–487.

Makinson, D. (1986). On the formal representation of rights relations: Remarks on the work of Stig Kanger and Lars Lindahl. *Journal of Philosophical Logic*, *15*, 403–425.

Prior, A. (1967). *Past, present, and future*. Oxford: Clarendon Press.

Reynolds, M. A. (2003). An axiomatization of Prior's Ockhamist logic of historical necessity. In P. Balbiani, N.-Y. Suzuki, F. Wolter, & M. Zakharyaschev (Eds.), *Advances in Modal Logic* (Vol. 4, pp. 355–370). London: King's College.

Schwarzentruber, F. (2012). Complexity results of STIT fragments. *Studia Logica*, *100*, 1001–1045.

Singh, M. P. (1999). An ontology for commitments in multiagent systems. *Artificial Intelligence and Law*, *7*, 97–113.

Singh, M. P. (2008). Semantic considerations on dialectical and practical commitments. In D. Fox & C. P. Gomes (Eds.), *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13–17, 2008* (pp. 176–181). Menlo Park, CA: AAAI Press.

Thomason, R. (1984). Combinations of tense and modality. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of Philosophical Logic* (2nd ed., Vol. 2, pp. 135–165). Berlin: Springer.

Vakarelov, D. (1992). A modal theory of arrows: Arrow logics I. In D. Pearce & G. Wagner (Eds.), *Logics in AI, European Workshop, JELIA '92, Berlin, Germany, September 7–10, 1992, proceedings* (pp. 1–24). Berlin: Springer.

von Kutschera, F. (1997). T x W completeness. *Journal of Philosophical Logic*, *26*, 241–250.

Wölf, S. (2002). Propositional Q-logic. *Journal of Philosophical Logic*, *31*, 387–414.

Xu, M. (1998). Axioms for deliberative STIT. *Journal of Philosophical Logic*, *27*, 505–552.

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