A Logical Analysis of Commitment Dynamics

Emiliano Lorini

▶ **To cite this version:**

Emiliano Lorini. A Logical Analysis of Commitment Dynamics. 10th International Conference on Deontic Logic in Computer Science (DEON 2010), Jul 2010, Florence, Italy. pp.288-305. hal-03672504

HAL Id: hal-03672504
https://hal.science/hal-03672504

Submitted on 30 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Logical Analysis of Commitment Dynamics

Emiliano Lorini

Université de Toulouse, IRIT-CNRS, France

Abstract. The aim of this work is to propose a model-theoretic semantics and a complete logic for the dynamics of commitments. In the first part of the article, a formalization of the concept of social commitment in STIT logic is presented. STIT is one of the most prominent formal accounts of agency. It is the logic of constructions of the form "agent i sees to it that φ". In the second part, the article presents an extension of STIT logic by dynamic operators which enable to describe two basic operations on commitment: commitment creation and commitment cancelation. The logic is used to develop an axiomatic and semantic analysis of commitment change in multi-agent systems.

1 Introduction

Social commitment is a fundamental concept for understanding normative relationships between individuals in a society. It has become a valuable abstraction for the design of multi-agent systems since it can be used to model a variety of interactive situations like contracts, agreements, negotiation, dialogue, and argumentation.

Although formal analysis of commitment are available in the literature (see, e.g., [27, 23, 6]) and commitments have been extensively used for applications in the area of multi-agent systems (see, e.g., [11] for an application of commitment to business protocols), there is still no formal approach which provides at the same time a *model-theoretic semantics*, and a *sound and complete logic* for the dynamics of commitments. I agree indeed with Singh [23, pp. 176] when he says that "...it was a sensible research strategy to first establish that commitments were a useful concept. However, now that the case for commitments has been made well, further progress is hampered by the lack of a clear model-theoretic semantics."

The aim of this article is to fill this existing gap in the literature on commitments by providing a model-theoretic semantics, and sound and complete logic for the dynamics of commitments. In order to formalize commitments STIT logic is used. STIT (the logic of *Seeing to it That*) is a logic of agency that has been developed in the 90ies in the domain of philosophy of action by Belnap, Horty and colleagues (see, e.g., [5, 15]). It is the logic of constructions of the form "agent i sees to it that φ".

In this article STIT logic is extended by modal operators which enable to express what is true according to the regulation of a given institution. With STIT logic augmented with these modal operators, I define commitments in institutional contexts. In particular, I define the concept of 'agent i's commitment towards agent j in the context of institution x to ensure a certain state of affairs φ' by the fact that, according to the regulation of institution x, if agent i does not see to it that φ then then agent j will be wronged by i.

The rest of the article is organized as follows. In Section 2, the STIT-based logical framework for the analysis of commitments is introduced. A formalization of the concept of social commitment is given in Section 3. In Section 4 an extension of the logic of Section 2 by dynamic operators is presented. These dynamic operators enable to describe two basic operations on commitment: *commitment creation* and *commitment cancelation*. The proposed logic is used to develop an axiomatic and semantic analysis of commitment change in multi-agent systems.

2 A logic of actions and institutions

In order to be able to reason about actions of agents and about pragmatic commitments in institutional contexts, I introduce here STIT logic extended by modal operators which enable to express what is true according to the regulation of a given institution. STIT logic (the logic of *Seeing to it That*) [5, 15] is one of the most prominent formal accounts of agency. It is the logic of constructions of the form "agent i sees to it that φ". STIT has a non-standard semantics based on the concepts of *moment* and *history*. However, as shown by Balbiani et al. [2], the basic STIT language without temporal operators axiomatized by Xu in [28] and [5, Chap. 17] can be 'simulated' in a standard Kripke semantics. Similarly to Balbiani et al., I use here a Kripke semantics for interpreting STIT modal operators.

Xu mainly focuses on Chellas's STIT operators named after his proponent [9]. As pointed out in [28, 15], so-called deliberative STIT operators and Chellas's STIT operators are interdefinable and just differ in the choice of primitive operators. Following Xu, I focus here on Chellas's STIT operators and I take them as primitive.

2.1 Syntax

Assume a countable set of atomic propositions denoting facts $Atm = \{p, q, \ldots\}$, a finite set of agents $Agt = \{i_1, \ldots, i_{|Agt|}\}$ and a finite set of institutional contexts $Inst = \{x_1, \ldots, x_{|Inst|}\}$. In order to be able to define social commitment in Section 3, let me add to the language special atoms as in the reduction of deontic logic to alethic logic [1]. In particular, let me introduce atoms of the form $\mathrm{wr}_{i,j,x}$ in Lindhal's style [19] one for every $i, j \in Agt$ such that $i \neq j$ and $x \in Inst$. I call Wr the corresponding set, that is,

$$Wr = \{\mathrm{wr}_{i,j,x} | i, j \in Agt \text{ and } i \neq j \text{ and } x \in Inst\}.$$

In the semantics, special atoms $\mathrm{wr}_{i,j,x}$ are used to identify those states in which, according to the regulation of a given institution x, an agent i wrongs (or gives offence to) another agent j. An atom $\mathrm{wr}_{i,j,x}$ has to be read 'agent j is wronged by agent i in the context of institution x'. I denote $Atm^+ = Atm \cup Wr$ the extended set of propositional atoms. I write $\alpha, \beta, \ldots$ the elements in Atm^+.

The language $\mathcal{L}$ of the logic $\mathbf{L}$ is the set of formulas defined by the following BNF:

$$\varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid [i]\varphi \mid \Box\varphi \mid \mathbf{D}_x\varphi$$

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

Operators $[i]$ are Chellas's STIT operators. Formula $[i]\varphi$ captures the fact that φ is guaranteed by a present choice 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'. I define the dual of the operator $[i]$ as follows: $\langle i\rangle\varphi \stackrel{\text{def}}{=} \neg[i]\neg\varphi$.

$\Box\varphi$ stands for 'φ is settled true regardless of what every agent does' or simply 'φ is settled 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 in [15]: $[i\ dstit{:}\ \varphi] \stackrel{\text{def}}{=} [i]\varphi \wedge \neg\Box\varphi$.

Finally, modal operators $\mathbf{D}_x$ enable to describe what is true according to the regulation of a given institution. In particular, formula $\mathbf{D}_x\varphi$ has the following reading 'φ is true, according to the regulation of the institution x' or more simply 'φ is true, according to the institution x'. For example, if the meaning of formula φ is that two persons are married and x denotes a certain State, formula $\mathbf{D}_x\varphi$ means that, according to the regulation of this State, these persons are considered as two legally married persons. However, it may happen that with respect to the regulation of another State y they are not married. I define the dual of the operator $\mathbf{D}_x$ as follows: $\widehat{\mathbf{D}}_x\varphi \stackrel{\text{def}}{=} \neg\mathbf{D}_x\neg\varphi$. Formula $\widehat{\mathbf{D}}_x\varphi$ has the following reading 'the regulation of the institution x admits a situation in which φ is true' or more simply 'the regulation of the institution x admits φ'.

This kind of modal operators were introduced for the first time by Jones & Sergot [16]. Similar operators were recently studied by Grossi et al. [13]. It has to be noted that Jones & Sergot's operators and Grossi et al.'s operators have slightly different properties. While the former satisfy the system KD and are interpreted by means of standard accessibility relations between worlds in a model, the latter are K45 operators which are interpreted by taking subsets of the set of worlds in a model. In this article, I adopt Jones & Sergot's solution by supposing that every $\mathbf{D}_x$ is a KD operator and by interpreting it by means of standard accessibility relations (see Section 2.3 below).

2.2 Semantics

I use a standard possible worlds semantics. Possible worlds are understood as in the logics of knowledge and of belief.

Definition 1 (L-model). ***L**-models are tuples $M = \langle W, \mathcal{R}_\Box, \{\mathcal{R}_i | i \in Agt\}, \{\mathcal{D}_x | x \in Inst\}, \mathcal{V}\rangle$ where:*

- *W is a nonempty set of possible worlds or states;*
- *$\mathcal{R}_\Box$ is an equivalence relation between worlds in W;*
- *for every $i \in Agt$, $\mathcal{R}_i$ is an equivalence relations between worlds in W such that:*
  (C1) *$\mathcal{R}_i \subseteq \mathcal{R}_\Box$,*
  (C2) *for all $w \in W$, for all $(w_j)_{j\in Agt} \in \mathcal{R}_\Box(w)^n$, $\bigcap_{j\in Agt} \mathcal{R}_j(w_j) \neq \emptyset$;*
- *for every $x \in Inst$, $\mathcal{D}_x$ is a serial relation between worlds in W such that:*
  (C3) *if $(w, v) \in \mathcal{D}_x$ and $(v, u) \in \mathcal{R}_\Box$ then $(w, u) \in \mathcal{D}_x$;*
- *$\mathcal{V} : Atm^+ \longrightarrow 2^W$ is a valuation function.*

For every $w \in W$, I write $|w| = \{\alpha \in Atm^+ | w \in \mathcal{V}(\alpha)\}$.

As in the previous Constraint C2, accessibility relations on W can be viewed as functions from W to 2^W. Therefore, I write $\mathcal{R}_i(w) = \{v|(w,v) \in \mathcal{R}_i\}$, $\mathcal{R}_\Box(w) = \{v|(w,v) \in \mathcal{R}_\Box\}$ and $\mathcal{D}_x(w) = \{v|(w,v) \in \mathcal{D}_x\}$. For every world $w \in W$ and for every agent $i \in Agt$, $\mathcal{R}_i(w)$ is the set of worlds that agent i brings about at world w or the set of outcomes of the action chosen by agent i at w. $\mathcal{R}_i(w)$ can also be called the *outcome state* of agent i at world w. $\mathcal{R}_\Box$ is the relation over all possible outcomes. If $v \in \mathcal{R}_\Box(w)$ then v is a possible outcome at w. $\mathcal{R}_\Box(w)$ is therefore called the *set of possible outcomes* at world w. If $v \in \mathcal{R}_\Box(w)$, we can also say that v is a possible alternative of world w.

Thus, Constraint C1 in Definition 1 just means that all outcomes brought about by an agent i are possible outcomes. Constraint C2 expresses a so-called *assumption of independence of agents*: if $w_1, \ldots, w_n$ are possible outcomes at w then the intersection of the set of outcomes that agent 1 brings about at w_1, and the set of outcomes that agent 2 brings about at w_2,..., and the set of outcomes that agent n brings about at w_n is not empty. More intuitively, this means that agents can never be deprived of choices due to the choices made by other agents.

Just as in epistemic logic an information state of an agent i is the set of worlds that agent i considers possible, the set $\mathcal{D}_x(w)$ is the *regulation state* of institution x at world w, that is, the set of worlds which are admitted by institution x's regulation at w. For example, suppose that x is a certain State and world v corresponds to the situation in which a couple $\{i_1, i_2\}$ applies for a divorce in x, and i_1 and i_2 are x's citizens, that is, $\{citizens_{i_1,i_2,x}, divorce_{i_1,i_2,x}\} \subseteq |v|$. Then, $v \in \mathcal{D}_x(w)$ just means that world v, in which a couple $\{i_1, i_2\}$ applies for a divorce in x and i_1 and i_2 are x's citizens, is admitted by State x's regulation at w.

Hence, Constraint C3 in Definition 1 just means that: if v is in the regulation state of an institution x at a world w then, all possible alternatives of v are worlds which are in the regulation state of institution x at world w.

Given a model M, a world w and a formula φ, we write $M, w \models \varphi$ to mean that φ is true at world w in M. The truth conditions of formulas are defined as follows:

- $M, w \models \alpha$ iff $w \in \mathcal{V}(\alpha)$;
- $M, w \models \neg\varphi$ iff not $M, w \models \varphi$;
- $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$;
- $M, w \models \Box\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{R}_\Box(w)$;
- $M, w \models [i]\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{R}_i(w)$;
- $M, w \models \mathbf{D}_x\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{D}_x(w)$.

I write $\models_{\mathbf{L}} \varphi$ if φ is *valid* in $\mathbf{L}$ (φ is true in all $\mathbf{L}$-models).

2.3 Axiomatization

Fig. 1 contains a complete axiomatization of the logic $\mathbf{L}$. We have all all principles of the normal modal logic S5 for every operator $[i]$ and for the operator $\Box$, and all principles of the normal modal logic KD for every operator $\mathbf{D}_x$. **($\Box \rightarrow i$)** and **($\mathbf{AIA}_k$)** are the two central principles in Xu's axiomatization of the Chellas's STIT operators $[i]$

PC	All principles of classical propositional calculus
S5(i)	All S5-principles for the operators $[i]$
S5($\Box$)	All S5-principles for the operator $\Box$
KD(x)	All KD-principles for the operators $\mathbf{D}_x$
($\Box \to i$)	$\Box\varphi \to [i]\varphi$
(AIA$_k$)	$(\Diamond[1]\varphi_1 \wedge \ldots \wedge \Diamond[k]\varphi_k) \to \Diamond([1]\varphi_1 \wedge \ldots \wedge [k]\varphi_k)$
($x \to x\Box$)	$\mathbf{D}_x\varphi \to \mathbf{D}_x\Box\varphi$

Fig. 1. Axiomatization of **L**

[28]. According to Axiom **($\Box \to i$)**, if φ is settled true then every agent sees to it that φ. In other words, an agent brings about those facts that are inevitable. **(AIA$_k$)** is a family of axiom schemes for independence of agents that is parameterized by the integer k. As noted in [5], **(AIA$_{k+1}$)** implies **(AIA$_k$)**. Therefore, as Agt is finite, the family of axiom schemas can be replaced by the single **(AIA$_{|Agt|}$)**. Finally, Axiom **($x \to x\Box$)** relates the modal operator $\Box$ with the institution operators $\mathbf{D}_x$. It says that if φ holds according to the institution x then, according to the institution x, φ is settled true.

I call **L** the logic axiomatized by the principles given in Fig. 1. I write $\vdash_{\mathbf{L}} \varphi$ if φ is a **L**-theorem.

Theorem 1. *The logic **L** is completely axiomatized by the principles in Fig. 1.*

Proof. It is a routine task to check that the axioms of the logic **L** correspond one-to-one to their semantic counterparts on the models. In particular, **S5**($\Box$) and **S5**(i) correspond to the fact that $\mathcal{R}_\Box$ and every $\mathcal{R}_i$ are equivalence relations, while **KD**(x) corresponds to the seriality of every $\mathcal{D}_x$. Axiom **($\Box \to i$)** corresponds to the Constraint C1 in Definition 1, while Axiom **($x \to x\Box$)** corresponds to the Constraint C3. Finally, as noted in Section 2.3, since Agt is finite, the family of axiom schemas **(AIA$_k$)** can be replaced by the single **(AIA$_{|Agt|}$)**. The latter corresponds to the Constraint C2 in Definition 1.

It is routine, too, to check that all axioms of the logic **L** are in the Sahlqvist class. This means that the axioms are all expressible as first-order conditions on models and that they are complete with respect to the defined model classes, cf. [7, Th. 2.42].

3 Commitments in institutional contexts: a formalization

According to [22, 8] a social commitment is a kind of normative relationship between a *debtor* and a *creditor* in a given *context*. 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 takes a pragmatic commitment towards j. On the contrary, if i tells to j: "Tomorrow, will be sunny. I am sure!" then, he takes a propositional commitment towards j.

In order to define the concept of social commitment, I use the special atoms $\mathrm{wr}_{i,j,x}$ denoting that 'an agent j is wronged by another agent i in the context of institution x'. I say that agent i is committed to agent j in the context of institution x to ensure φ (noted $\mathbf{C}_{i:j:x}\varphi$) if and only if, according to the institution x if i does not see to it that φ then j will be wronged by i in x, and the institution x admits a situation in which i does not see to it that φ. For every $i, j \in Agt$ and $x \in Inst$ I define:

$$\mathbf{C}_{i:j:x}\varphi \stackrel{\text{def}}{=} \mathbf{D}_x(\neg[i]\varphi \rightarrow \mathrm{wr}_{i,j,x}) \wedge \widehat{\mathbf{D}}_x\neg[i]\varphi.$$

The component $\widehat{\mathbf{D}}_x\neg[i]\varphi$ expresses that i is committed to j in the context of institution x to ensure φ, only if the situation in which i does not see to it that φ (and j is therefore wronged by j) is compatible with the institution x's regulation. It has also to be noted that a commitment of agent i to agent j to ensure φ in the context of institution x can be conceived as a kind of *directed obligation* from a bearer to a counterparty in a given institutional context (see, e.g., [12, 14, 17, 19, 21] for some analysis of the notion of directed obligation in deontic logic).[1]

Example 1. Agent i_2 is the program chair of a given conference. Agent i_2 asks agent i_1, a member of the program committee, to review some articles submitted to the conference. Agent i_1 accepts agent's i_2 request by sending a confirmation e-mail (we suppose that the communication between i_1 and i_2 is made through the Easychair system). Consequently, according to the program committee of the conference, agent i_1 is committed to i_2 to review the articles: $\mathbf{C}_{i_1:i_2:PC}\mathit{review}$. This means that, according to the program committee, if i_1 does not review the articles then i_2 will be wronged by i_1. Moreover, the program committee admits a situation in which i_1 does not accomplish his duty to review the articles:

$$\mathbf{D}_{PC}(\neg[i_1]\mathit{review} \rightarrow \mathrm{wr}_{i_1,i_2,PC}) \wedge \widehat{\mathbf{D}}_{PC}\neg[i_1]\mathit{review}.$$

Let me generalize the previous definition to conditional commitments. I say that agent i is committed to agent j in the context of institution x to ensure φ under condition ψ (noted $\mathbf{C}_{i:j:x}(\psi,\varphi)$) if and only if, according to the institution x, if i does not see to it that φ and ψ is true then j will be wronged by i, and the institution x admits a situation in which ψ is true and agent i does not see to it that φ. For every $i, j \in Agt$ and $x \in Inst$ I define:

$$\mathbf{C}_{i:j:x}(\psi,\varphi) \stackrel{\text{def}}{=} \mathbf{D}_x((\psi \wedge \neg[i]\varphi) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \widehat{\mathbf{D}}_x(\psi \wedge \neg[i]\varphi).$$

Again the formula $\widehat{\mathbf{D}}_x(\psi \wedge \neg[i]\varphi)$ expresses that i is committed to j in the context of institution x to ensure φ under condition ψ, only if the situation in which ψ is true and i does not see to it that φ is compatible with the institution x's regulation. Note also

[1] In legal theory it is typically assumed that a directed obligation for i towards j to ensure φ correlates to a *right* for j towards i that φ is brought about by i.

that $\mathbf{D}_x\neg\psi$ implies $\neg\mathbf{C}_{i:j:x}(\psi,\varphi)$ for every formula φ and for every couple of agents i and j. That is, if the institution x does not admit a situation in which ψ is true then, for every formula φ and for every couple of agents i and j, there is no commitment of i towards j in the context of institution x to ensure φ under condition ψ.

Example 2. Agent i_1 and agent i_2 have concluded a contract in front of a notary of a certain State x. Agent i_1 has declared in front of the notary that he will sell to i_2 a certain property if i_2 will pay him 10K Euros. Consequently, according to the State x, i_1 is conditionally committed to i_2 to sell to i_2 the property under the condition that i_2 pays him 10K Euros:

$$\mathbf{C}_{i_1:i_2:x}([i_2]pay10K_{i_1},sellProperty_{i_2}).$$

This means that, according to the State x, if i_1 does not sell to i_2 the property when i_2 pays him 10K Euros then i_2 will be wronged by i_2, and the State x admits a situation in which i_1 does not sell to i_2 the property while i_2 pays him 10K Euros:

$$\mathbf{D}_x(([i_2]pay10K_{i_1}\wedge\neg[i_1]sellProperty_{i_2})\rightarrow \mathrm{wr}_{i_1,i_2,x})\wedge \widehat{\mathbf{D}}_x([i_2]pay10K_{i_1}\wedge\neg[i_1]sellProperty_{i_2}).$$

As the following $\mathbf{L}$-theorem highlights in the present approach unconditional commitments are special cases of conditional commitments where the antecedent is true:

(1) $$\vdash_{\mathbf{L}} \mathbf{C}_{i:j:x}\varphi\leftrightarrow\mathbf{C}_{i:j:x}(\top,\varphi).$$

It is also interesting to note that the previous definition satisfies some intuitive properties of conditional commitments. For every $i,j\in Agt$ and for every $x\in Inst$ we have:

(2) $$\vdash_{\mathbf{L}} \mathbf{D}_x\varphi\rightarrow\neg\mathbf{C}_{i:j:x}(\psi,\varphi)$$

(3) $$\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{D}_x\psi)\rightarrow\mathbf{C}_{i:j:x}\varphi$$

(4) $$\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{C}_{i:j:x}(\chi,\varphi))\rightarrow\mathbf{C}_{i:j:x}(\psi\vee\chi,\varphi)$$

(5) $$\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{C}_{i:j:x}(\psi,\chi))\rightarrow\mathbf{C}_{i:j:x}(\psi,\varphi\wedge\chi)$$

(6) $$\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi\wedge\chi)\wedge\mathbf{D}_x\chi)\rightarrow\mathbf{C}_{i:j:x}(\psi,\varphi)$$

Proof. We prove $\mathbf{L}$-theorem 5 as an example.

1. $\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{C}_{i:j:x}(\psi,\chi))\leftrightarrow$
   $(\mathbf{D}_x((\psi\wedge\neg[i]\varphi)\rightarrow\mathrm{wr}_{i,j,x})\wedge\mathbf{D}_x((\psi\wedge\neg[i]\chi)\rightarrow\mathrm{wr}_{i,j,x}))$
2. $\vdash_{\mathbf{L}} (\mathbf{D}_x((\psi\wedge\neg[i]\varphi)\rightarrow\mathrm{wr}_{i,j,x})\wedge\mathbf{D}_x((\psi\wedge\neg[i]\chi)\rightarrow\mathrm{wr}_{i,j,x}))\rightarrow$
   $\mathbf{D}_x((\psi\wedge(\neg[i]\varphi\vee\neg[i]\chi))\rightarrow\mathrm{wr}_{i,j,x})$ by standard modal principles for $\mathbf{D}_x$
3. $\vdash_{\mathbf{L}} \mathbf{D}_x((\psi\wedge(\neg[i]\varphi\vee\neg[i]\chi))\rightarrow\mathrm{wr}_{i,j,x})\rightarrow\mathbf{D}_x((\psi\wedge\neg[i](\varphi\wedge\chi))\rightarrow\mathrm{wr}_{i,j,x})$
   by standard modal principles for $[i]$ and $\mathbf{D}_x$, and necessitation for $\mathbf{D}_x$
4. $\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{C}_{i:j:x}(\psi,\chi))\rightarrow(\widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle\neg\varphi)\wedge\widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle\neg\chi))$
5. $\vdash_{\mathbf{L}} (\widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle\neg\varphi)\wedge\widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle\neg\chi))\rightarrow\widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle(\neg\varphi\vee\neg\chi))$
   by standard modal principles for $[i]$ and $\mathbf{D}_x$
6. $\vdash_{\mathbf{L}} \widehat{\mathbf{D}}_x(\psi\wedge\langle i\rangle(\neg\varphi\vee\neg\chi))\rightarrow\widehat{\mathbf{D}}_x(\psi\wedge\neg[i](\varphi\wedge\chi))$
7. $\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi)\wedge\mathbf{C}_{i:j:x}(\psi,\chi))\rightarrow$
   $(\mathbf{D}_x((\psi\wedge\neg[i](\varphi\wedge\chi))\rightarrow\mathrm{wr}_{i,j,x})\wedge\widehat{\mathbf{D}}_x(\psi\wedge\neg[i](\varphi\wedge\chi)))$ from 1-3 and 4-6

8. $\vdash_{\mathbf{L}} (\mathbf{C}_{i:j:x}(\psi,\varphi) \wedge \mathbf{C}_{i:j:x}(\psi,\chi)) \rightarrow \mathbf{C}_{i:j:x}(\psi,\varphi \wedge \chi)$ from 7

L-theorem 2 is a *discharge* principle for commitment: if according to institution x φ is true, then i's commitment towards j in x to ensure φ is discharged and is no longer active. In the logic **L** $\mathbf{D}_x\varphi$ is equivalent to $\mathbf{D}_x[i]\varphi$ (by Axiom ($x \rightarrow x\Box$) and Axiom T for $[i]$). Therefore, **L**-theorem 2 can be written in the following equivalent form: $\mathbf{D}_x[i]\varphi \rightarrow \neg\mathbf{C}_{i:j:x}(\psi,\varphi)$. Note also that **L**-theorem 2 and the **L**-theorem $\mathbf{D}_x\top$ together imply $\neg\mathbf{C}_{i:j:x}(\psi,\top)$: an agent i cannot be committed to bring about tautologies.

L-theorem 3 is a *detachment* principle: if i is conditionally committed to j in x to ensure φ if ψ holds and, according to institution x, ψ holds then, an unconditional commitment of i to j in x to ensure φ comes into being. A similar detachment principle for conditional obligations has been discussed by Bartha [4].

L-theorems 4 and 5 are respectively a *disjunction of the antecedents* principle and a *conjunction of the consequents* principle for commitment. Similar properties have been isolated in [23]. According to **L**-theorem 4, if i is committed to j in x to ensure φ if ψ holds and to ensure φ if χ holds then, i is committed to j in x to ensure φ if $\psi \vee \chi$ holds. According to **L**-theorem 5, if i is committed to j in x to ensure φ if ψ holds and to ensure χ if ψ holds then, i is committed to j in x to ensure $\varphi \wedge \chi$ if ψ holds.

L-theorem 6 is a *weakening* principle for commitment: if agent i is committed to agent j in x to ensure $\varphi \wedge \chi$ if ψ holds and, according to the institution x, χ is true, then i is committed to j to ensure φ if ψ holds. So, if an agent is committed to ensure two states of affairs φ and χ and his commitment to ensure χ is discharged, then the agent is committed to ensure φ. Note indeed that, as highlighted by **L**-theorem 2, $\mathbf{D}_x\chi$ is the discharge condition for the commitment $\mathbf{C}_{i:j:x}(\psi,\chi)$.

Before concluding, let me consider how the previous definition of commitment behaves in the case of Moore-like sentences of the form $\varphi \wedge \neg[i]\varphi$. The following **L**-theorem 7 clarifies this point: an agent is committed to ensure that φ is true and that he does not see to it that φ if and only if, the agent is committed to do something inconsistent.

(7) $$\vdash_{\mathbf{L}} \mathbf{C}_{i:j:x}(\varphi \wedge \neg[i]\varphi) \leftrightarrow \mathbf{C}_{i:j:x}\bot.$$

4 From static to dynamic commitments

I here extend the logic **L** of Section 1 by dynamic operators which enable to describe the dynamics of social commitments. I call $\mathbf{L}^{dyn}$ the extended logic. I consider two basic operations on commitment: commitment creation and commitment cancelation. These two kinds of operations have also been studied in [22] and [27] in which commitments are operationalized as being in a certain number of states and operations of commitment creation and commitment cancelation are responsible for changing the state of a commitment from *passive* to *active* and viceversa. The main contribution of this section is to provide a comprehensive logical approach to the dynamics of commitments and, in particular, a model-theoretic semantics and a complete modal logic for commitment dynamics.

Operations of commitment creation are of the form $i{:}j{:}x{+}(\psi,\varphi)$, operations of commitment cancelation are of the form $i{:}j{:}x{-}(\psi,\varphi)$. In particular, $i{:}j{:}x{+}(\psi,\varphi)$ is the

event 'the commitment of agent i towards agent j to ensure φ if ψ holds is created in the institution x' whereas $i{:}j{:}x{-}(\psi,\varphi)$ is the event 'the commitment of agent i towards agent j to ensure φ if ψ holds is canceled from the institution x'.

4.1 Commitment creation

The first extension of the logic **L** is by formulas $[i{:}j{:}x{+}(\psi,\varphi)]\chi$ where i, j ranges over Agt and x ranges over $Inst$. Formula $[i{:}j{:}x{+}(\psi,\varphi)]\chi$ describes the effects of the creation in institution x of i's commitment towards j to ensure φ if ψ holds. More precisely, $[i{:}j{:}x{+}(\psi,\varphi)]\chi$ has to be read 'χ holds, after the creation in the institution x of agent i's commitment towards agent j to ensure φ if ψ holds'.

Semantics In order to give semantics to the operators $[i{:}j{:}x{+}(\psi,\varphi)]$ I need to define the model $M^{i:j:x+(\psi,\varphi)}$ which results from the occurrence of the event $i{:}j{:}x{+}(\psi,\varphi)$ in the model M. The elements of the model $M^{i:j:x+(\psi,\varphi)}$ are defined as follows:

Definition 2 (Updated model $M^{i:j:x+(\psi,\varphi)}$). *For every **L**-model M, $M^{i:j:x+(\psi,\varphi)}$ is the corresponding updated model with:*

$$
\begin{aligned}
W^{i:j:x+(\psi,\varphi)} =& \{w_x | w \in W\} \cup \{w_{\sim x} | w \in W\};\\
\mathcal{D}_x^{i:j:x+(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{D}_x\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{D}_x\};\\
\textit{If } y \neq x \textit{ then, } \mathcal{D}_y^{i:j:x+(\psi,\varphi)} =& \{(w_x, v_{\sim x}) | (w,v) \in \mathcal{D}_y\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{D}_y\};\\
\mathcal{R}_\Box^{i:j:x+(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{R}_\Box\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{R}_\Box\};\\
\textit{For } z \in \textit{Agt, } \mathcal{R}_z^{i:j:x+(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{R}_z\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{R}_z\};\\
\mathcal{V}^{i:j:x+(\psi,\varphi)}(\mathrm{wr}_{i,j,x}) =& \{w_x | M, w \models \mathrm{wr}_{i,j,x} \vee (\psi \wedge \neg[i]\varphi)\} \cup\\
& \{w_{\sim x} | M, w \models \mathrm{wr}_{i,j,x}\};\\
\textit{For } \alpha \neq \mathrm{wr}_{i,j,x}, \mathcal{V}^{i:j:x+(\psi,\varphi)}(\alpha) =& \{w_x | M, w \models \alpha\} \cup \{w_{\sim x} | M, w \models \alpha\}.
\end{aligned}
$$

$M^{i:j:x+(\psi,\varphi)}$ is obtained by creating two copies of each world w of the original model M: a copy w_x for the regulation state of institution x, also called 'x-copy' of world w; a copy $w_{\sim x}$ for the regulation states of every institution y different from x, also called '$\sim x$-copy' of world w. World w_x is the copy of w which is affected by the occurrence of the normative event $i{:}j{:}x{+}(\psi,\varphi)$, whereas $w_{\sim x}$ is the copy of w in which nothing changes. In particular, in w_x the truth value of the atom $\mathrm{wr}_{i,j,x}$ ('agent j is wronged by agent i in x') is set to true if and only if, $\mathrm{wr}_{i,j,x}$ is already true at w, or at w ψ is true and i does not see to it that φ (i.e. $\psi \wedge \neg[i]\varphi$). Moreover, for every copy w_x and for every copy $w_{\sim x}$:

- the regulation state of institution x at w_x are the 'x-copies' of worlds that were in the regulation state of x at w before the event $i{:}j{:}x{+}(\psi,\varphi)$;
- the regulation state of every institution y different from x at w_x are the '$\sim x$-copies' of worlds that were in the regulation state of y at w before the event $i{:}j{:}x{+}(\psi,\varphi)$;
- the regulation state of every institution y at $w_{\sim x}$ are the '$\sim x$-copies' of worlds that were in the regulation state of y at w before the event $i{:}j{:}x{+}(\psi,\varphi)$;

– the outcome state of every agent i (resp. the set of possible states) at w_x are the 'x-copies' of worlds that were in the outcome state of i (resp. in the set of possible states) at w before the event $i{:}j{:}x{+}(\psi,\varphi)$;
– the outcome state of every agent i (resp. the set of possible states) at $w_{\sim x}$ are the '$\sim x$-copies' of worlds that were in the outcome state of i (resp. in the set of possible states) at w before the event $i{:}j{:}x{+}(\psi,\varphi)$.

As the following proposition highlights the operation of commitment creation is well-defined.

Proposition 1. *If M is a **L**-model then $M^{i:j:x+(\psi,\varphi)}$ is a **L**-model.*

Proof. It is just straightforward to show that $\mathcal{R}_{\Box}^{i:j:x+(\psi,\varphi)}$ and every $\mathcal{R}_{z}^{i:j:x+(\psi,\varphi)}$ (with $z \in Agt$) are equivalence relations, and that every $\mathcal{D}_{x}^{i:j:x+(\psi,\varphi)}$ (with $x \in Inst$) is a serial relation. It is also a routine to check that the operation $i{:}j{:}x{+}(\psi,\varphi)$ preserves Constraints C1 and C2 in Definition 1.

Let me prove that the operation $i{:}j{:}x{+}(\psi,\varphi)$ also preserves Constraint C3. Suppose that $(w_x, v_x) \in \mathcal{D}_{x}^{i:j:x+(\psi,\varphi)}$ and $(v_x, u_x) \in \mathcal{R}_{\Box}^{i:j:x+(\psi,\varphi)}$. Then, we have $(w, v) \in \mathcal{D}_x$ and $(v, u) \in \mathcal{R}_{\Box}$. By Constraint C3 on **L**-models, it follows that $(w, u) \in \mathcal{D}_x$. Consequently, we have $(w_x, u_x) \in \mathcal{D}_{x}^{i:j:x+(\psi,\varphi)}$. In a similar way we can prove that if $(w_{\sim x}, v_{\sim x}) \in \mathcal{D}_{x}^{i:j:x+(\psi,\varphi)}$ and $(v_{\sim x}, u_{\sim x}) \in \mathcal{R}_{\Box}^{i:j:x+(\psi,\varphi)}$ then $(w_{\sim x}, u_{\sim x}) \in \mathcal{D}_{x}^{i:j:x+(\psi,\varphi)}$.

Now consider the case $y \neq x$ and suppose that $(w_x, v_{\sim x}) \in \mathcal{D}_{y}^{i:j:x+(\psi,\varphi)}$ and $(v_{\sim x}, u_{\sim x}) \in \mathcal{R}_{\Box}^{i:j:x+(\psi,\varphi)}$. Then, we have $(w, v) \in \mathcal{D}_y$ and $(v, u) \in \mathcal{R}_{\Box}$. By Constraint C3 on **L**-models, it follows that $(w, u) \in \mathcal{D}_y$. Consequently, we have $(w_x, u_{\sim x}) \in \mathcal{D}_{y}^{i:j:x+(\psi,\varphi)}$. In a similar way we can prove that if $(w_{\sim x}, v_{\sim x}) \in \mathcal{D}_{y}^{i:j:x+(\psi,\varphi)}$ and $(v_{\sim x}, u_{\sim x}) \in \mathcal{R}_{\Box}^{i:j:x+(\psi,\varphi)}$ then $(w_{\sim x}, u_{\sim x}) \in \mathcal{D}_{y}^{i:j:x+(\psi,\varphi)}$.

The truth conditions of the operators $[i{:}j{:}x{+}(\psi,\varphi)]$ are the following:

$$M, w \models [i{:}j{:}x{+}(\psi,\varphi)]\chi \text{ iff } M^{i:j:x+(\psi,\varphi)}, w_x \models \chi.$$

Thus, at world w of model M it is the case that χ holds, after the creation in the institution x of i's commitment towards j to ensure φ if ψ holds if and only if, χ holds at the x-copy w_x of world w in the updated model $M^{i:j:x+(\psi,\varphi)}$.

Example 3. Fig. 2 illustrates by means of an example the semantics of the operation of commitment creation. Formulas $\mathbf{C}_{1:2:x}q$ and $\mathbf{C}_{1:2:y}q$ are true at world w in the initial model (at w agent 1 is committed to agent 2 to ensure q both in the context of institution x and in the context of institution y), whereas formulas $\mathbf{C}_{1:2:x}p$ and $\mathbf{C}_{1:2:y}p$ are both false at w (at w 1 is not committed to agent 2 to ensure p). The operation $1{:}2{:}x{+}(\top, p)$ results in an updated model in which 1 becomes committed to 2 in x to ensure p and in which 1 remains committed to 2 in x to ensure q. On the contrary, 1's commitments towards 2 in the context of institution y do not change. Indeed, $\mathbf{C}_{1:2:x}p$, $\mathbf{C}_{1:2:x}q$ and $\mathbf{C}_{1:2:y}q$ are true at world w_x in the updated model, whereas $\mathbf{C}_{1:2:y}p$ is false at w_x.

Note that the updated model is nothing else than the duplication of the initial model. The left copy is the 'x-copy' which is affected by the event $1{:}2{:}x{+}(\top, p)$, whereas the right copy is the '$\sim x$-copy' in which nothing changes.

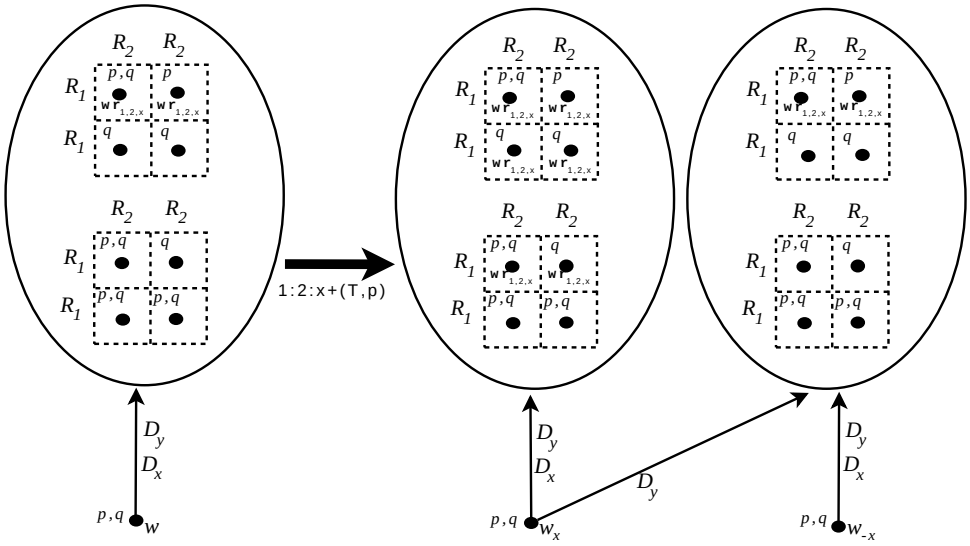

Fig. 2. Example of commitment creation. The left side and the right side of the picture respectively represent the model before the operation of commitment creation $1{:}2{:}x{+}(\top, p)$ and the model resulting from this operation (the updated model). Vertical circles in the two models represent the regulation states of institution x and of institution y. For instance in the initial model, institution x and institution y have the same regulation state at world w. Inside each regulation state, there are two sets of possible outcomes (the two dotted grids). The rows in each grid are the outcome states of agent 1, whereas the columns are the outcome states of agent 2.

Axiomatization The following lemma is fundamental in order to prove Theorem 2 below and can be easily proved by induction on the structure of χ.

Lemma 1. *Take a* ***L****-model* M*. Then, for every* $w_{\sim x} \in W^{i:j:x+(\psi,\varphi)}$*, we have:* $M^{i:j:x+(\psi,\varphi)}, w_{\sim x} \models \chi$ *iff* $M, w \models \chi$*.*

As the following theorem highlights, there are reduction axioms for the operators $[i{:}j{:}x{+}(\psi,\varphi)]$. They are called reduction axioms because, read from left to right, they reduce the complexity of those operators in a formula.

Theorem 2. *The following equivalences are valid:*

R1. $[i{:}j{:}x{+}(\psi,\varphi)]\mathrm{wr}_{i,j,x} \leftrightarrow (\mathrm{wr}_{i,j,x} \vee (\psi \wedge \neg[i]\varphi))$

R2. $[i{:}j{:}x{+}(\psi,\varphi)]\alpha \leftrightarrow \alpha$ *if* $\alpha \neq \mathrm{wr}_{i,j,x}$

R3. $[i{:}j{:}x{+}(\psi,\varphi)]\neg\chi \leftrightarrow \neg[i{:}j{:}x{+}(\psi,\varphi)]\chi$

R4. $[i{:}j{:}x{+}(\psi,\varphi)](\chi_1 \wedge \chi_2) \leftrightarrow ([i{:}j{:}x{+}(\psi,\varphi)]\chi_1 \wedge [i{:}j{:}x{+}(\psi,\varphi)]\chi_2)$

R5. $[i{:}j{:}x{+}(\psi,\varphi)]\mathbf{D}_x\chi \leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(\psi,\varphi)]\chi$

R6. $[i{:}j{:}x{+}(\psi,\varphi)]\mathbf{D}_y\chi \leftrightarrow \mathbf{D}_y\chi$ *if* $y \neq x$

R7. $[i{:}j{:}x{+}(\psi,\varphi)][i]\chi \leftrightarrow [i][i{:}j{:}x{+}(\psi,\varphi)]\chi$

R8. $[i{:}j{:}x{+}(\psi,\varphi)]\Box\chi \leftrightarrow \Box[i{:}j{:}x{+}(\psi,\varphi)]\chi$

Proof. We just prove **R5** and **R6** as examples.
(**R5**) $M, w \models [i{:}j{:}x{+}(\psi,\varphi)]\mathbf{D}_x\chi$,
$\Leftrightarrow M^{i:j:x+(\psi,\varphi)}, w_x \models \mathbf{D}_x\chi$,
$\Leftrightarrow$ for all $v_x \in \mathcal{D}_x(w_x)$, $M^{i:j:x+(\psi,\varphi)}, v_x \models \chi$,
$\Leftrightarrow$ for all $v \in \mathcal{D}_x(w)$, $M, v \models [i{:}j{:}x{+}(\psi,\varphi)]\chi$,
$\Leftrightarrow M, w \models \mathbf{D}_x[i{:}j{:}x{+}(\psi,\varphi)]\chi$.
(**R6**) Suppose $y \neq x$. Then:
$M, w \models [i{:}j{:}x{+}(\psi,\varphi)]\mathbf{D}_y\chi$,
$\Leftrightarrow M^{i:j:x+(\psi,\varphi)}, w_x \models \mathbf{D}_y\chi$,
$\Leftrightarrow$ for all $v_{\sim x} \in \mathcal{D}_x(w_x)$, $M^{i:j:x+(\psi,\varphi)}, v_{\sim x} \models \chi$,
$\Leftrightarrow$ for all $v \in \mathcal{D}_x(w)$, $M, v \models \chi$ (by Lemma 1),
$\Leftrightarrow M, w \models \mathbf{D}_x\chi$.

Some properties The following are examples of valid properties of commitment creation, with $i, j, h, k \in Agt$, $p, q \in Atm$, and $x, y \in Inst$:

(8) $\widehat{\mathbf{D}}_x(q \wedge \neg[i]p) \leftrightarrow [i{:}j{:}x{+}(q,p)]\mathbf{C}_{i:j:x}(q,p)$

(9) $\mathbf{C}_{h:k:x}(q,p) \rightarrow [i{:}j{:}x{+}(\psi,\varphi)]\mathbf{C}_{h:k:x}(q,p)$

(10) $\neg\mathbf{C}_{h:k:x}(q,p) \rightarrow [i{:}j{:}x{+}(\psi,\varphi)]\neg\mathbf{C}_{h:k:x}(q,p)$ if $h \neq i$ or $k \neq j$

(11) $\mathbf{C}_{h:k:y}(\gamma,\chi) \rightarrow [i{:}j{:}x{+}(\psi,\varphi)]\mathbf{C}_{h:k:y}(\gamma,\chi)$ if $y \neq x$

(12) $\neg\mathbf{C}_{h:k:y}(\gamma,\chi) \rightarrow [i{:}j{:}x{+}(\psi,\varphi)]\neg\mathbf{C}_{h:k:y}(\gamma,\chi)$ if $y \neq x$

Proof. We prove $\mathbf{L}^{dyn}$-theorem 8 as an example by applying the reduction axioms of Theorem 2 together with the rule of replacement of proved equivalence.
$[i{:}j{:}x{+}(q,p)]\mathbf{C}_{i:j:x}(q,p)$
$\Leftrightarrow [i{:}j{:}x{+}(q,p)]\mathbf{D}_x((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge [i{:}j{:}x{+}(q,p)]\neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R4**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x[i{:}j{:}x{+}(q,p)](q \rightarrow [i]p)$
(by **R3**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x\neg[i{:}j{:}x{+}(q,p)](q \wedge \neg[i]p)$
(by **R3**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge$
$\neg\mathbf{D}_x\neg([i{:}j{:}x{+}(q,p)]q \wedge [i{:}j{:}x{+}(q,p)]\neg[i]p)$
(by **R4**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x\neg(q \wedge \neg[i{:}j{:}x{+}(q,p)][i]p)$
(by **R2** and **R3**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x\neg(q \wedge \neg[i][i{:}j{:}x{+}(q,p)]p)$
(by **R7**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x\neg(q \wedge \neg[i]p)$
(by **R2**),
$\Leftrightarrow \mathbf{D}_x[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \rightarrow \mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$,
$\Leftrightarrow \mathbf{D}_x\neg[i{:}j{:}x{+}(q,p)]((q \wedge \neg[i]p) \wedge \neg\mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R3**),
$\Leftrightarrow \mathbf{D}_x\neg([i{:}j{:}x{+}(q,p)](q \wedge \neg[i]p) \wedge [i{:}j{:}x{+}(q,p)]\neg\mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R4**),

$\Leftrightarrow \mathbf{D}_x([i{:}j{:}x{+}(q,p)](q \rightarrow [i]p) \vee [i{:}j{:}x{+}(q,p)]\mathrm{wr}_{i,j,x}) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R3** and the valid equivalence $[i{:}j{:}x{+}(q,p)]\varphi \leftrightarrow \neg[i{:}j{:}x{+}(q,p)]\neg\varphi$),
$\Leftrightarrow \mathbf{D}_x([i{:}j{:}x{+}(q,p)](q \rightarrow [i]p) \vee \mathrm{wr}_{i,j,x} \vee (q \wedge \neg[i]p)) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R7**),
$\Leftrightarrow \mathbf{D}_x((q \rightarrow [i]p) \vee \mathrm{wr}_{i,j,x} \vee (q \wedge \neg[i]p)) \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$
(by **R2**, **R3**, **R4**, **R7**),
$\Leftrightarrow \mathbf{D}_x\top \wedge \neg\mathbf{D}_x(q \rightarrow [i]p)$,
$\Leftrightarrow \neg\mathbf{D}_x(q \rightarrow [i]p)$,
$\Leftrightarrow \widehat{\mathbf{D}}_x(q \wedge \neg[i]p)$.

$\mathbf{L}^{dyn}$-theorem 8 highlights that the operator of commitment creation is well-defined since it captures the desired commitment dynamics. Suppose p and q are propositional atoms in Atm. If the institution x admits a situation in which q is true and agent i does not see to it that p then, after the creation in x of i's commitment towards j to ensure p if q holds, i will be committed to j in x to ensure p if q holds, and viceversa.

According to theorem 9, the operation of creating a commitment in a given institution x does not cancel pre-existing commitments about propositions in the same institution. In other words, the operation of commitment creation is conservative. Note that this property is due to the fact that in the model resulting from the operation of commitment creation, if $\mathrm{wr}_{i,j,x}$ is true at world w in the original model then $\mathrm{wr}_{i,j,x}$ is set to true at w_x in the updated model. Consequently, the set of worlds in the 'x-copy' of the original model in which $\mathrm{wr}_{i,j,x}$ is true is larger than the set of worlds in the original model in which $\mathrm{wr}_{i,j,x}$ is true. According to theorem 10, the operation of creating a new commitment about propositions from a debtor i to a creditor j in a given institution x does not create additional commitments in the same institution from a debtor h to a creditor k, if either h and i are different debtors or k and j are different creditors. Finally, theorems 11 and 12 highlight some locality aspects of commitment creation, where locality means that the process of creating a commitment in an institution x does not change the commitments in an institution different from x.

REMARK. Note that the following formula is also valid in the logic $\mathbf{L}^{dyn}$ for every $p, q \in Atm$:

$$[i{:}j{:}x{+}(q,p)]\Box(\neg[i](q \rightarrow [i]p) \rightarrow \mathrm{wr}_{i,j,x}).$$

This means, after the creation in institution x of agent i's commitment towards agent j to ensure p if q holds, it is settled that, if i does not see to it that if ψ then he sees to it that φ, then j will be wronged by i in x. In other words, the normative consequences of an operation of commitment creation are necessarily true facts.

4.2 Commitment cancelation

The second extension of the logic $\mathbf{L}$ is by formulas $[i{:}j{:}x{-}(\psi,\varphi)]\chi$, where i, j ranges over Agt and x ranges over $Inst$. Formula $[i{:}j{:}x{-}(\psi,\varphi)]\chi$ has to be read 'χ holds, after the cancelation from the institution x of agent i's commitment towards agent j to ensure φ if ψ holds'.

Semantics In order to give semantics to the operators $[i{:}j{:}x{-}(\psi,\varphi)]$ let me define the model $M^{i:j:x-(\psi,\varphi)}$ which results from the cancelation from the institution x of i's commitment towards j to ensure φ if ψ holds.

Definition 3 (Updated model $M^{i:j:x-(\psi,\varphi)}$). *For every **L**-model M, $M^{i:j:x-(\psi,\varphi)}$ is the corresponding updated model with:*

$$
\begin{aligned}
W^{i:j:x-(\psi,\varphi)} =& \{w_x | w \in W\} \cup \{w_{\sim x} | w \in W\};\\
\mathcal{D}_x^{i:j:x-(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{D}_x\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{D}_x\};\\
\textit{If } y \neq x \textit{ then, } \mathcal{D}_y^{i:j:x-(\psi,\varphi)} =& \{(w_x, v_{\sim x}) | (w,v) \in \mathcal{D}_y\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{D}_y\};\\
\mathcal{R}_\Box^{i:j:x-(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{R}_\Box\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{R}_\Box\};\\
\textit{For } z \in \textit{Agt, } \mathcal{R}_z^{i:j:x-(\psi,\varphi)} =& \{(w_x, v_x) | (w,v) \in \mathcal{R}_z\} \cup \{(w_{\sim x}, v_{\sim x}) | (w,v) \in \mathcal{R}_z\};\\
\mathcal{V}^{i:j:x-(\psi,\varphi)}(\mathrm{wr}_{i,j,x}) =& \{w_x | M, w \models \mathrm{wr}_{i,j,x} \wedge (\psi \rightarrow [i]\varphi)\} \cup\\
& \{w_{\sim x} | M, w \models \mathrm{wr}_{i,j,x}\};\\
\textit{For } \alpha \neq \mathrm{wr}_{i,j,x}, \mathcal{V}^{i:j:x-(\psi,\varphi)}(\alpha) =& \{w_x | M, w \models \alpha\} \cup \{w_{\sim x} | M, w \models \alpha\}.
\end{aligned}
$$

$M^{i:j:x-(\psi,\varphi)}$ is also obtained by creating two copies of each world w of the original model M. Again, world w_x (the 'x-copy') is the copy of w which is affected by the occurrence of the event $i{:}j{:}x{-}(\psi,\varphi)$, whereas $w_{\sim x}$ (the '$\sim x$-copy') is the copy in which nothing happens. In particular, in w_x the truth value of the atom $\mathrm{wr}_{i,j,x}$ is set to true if and only if, at w $\mathrm{wr}_{i,j,x}$ is true and if ψ is true then i sees to it that φ (i.e. $\psi \rightarrow [i]\varphi$). For the rest, model $M^{i:j:x-(\psi,\varphi)}$ has the same structure as the model $M^{i:j:x+(\psi,\varphi)}$ defined in Section 4.1.

Proposition 2. *If M is a **L**-model then $M^{i:j:x-(\psi,\varphi)}$ is a **L**-model.*

The truth condition of $[i{:}j{:}x{-}(\psi,\varphi)]\chi$ is the following:

$$M, w \models [i{:}j{:}x{-}(\psi,\varphi)]\chi \text{ iff } M^{i:j:x-(\psi,\varphi)}, w_x \models \chi.$$

Axiomatization The following Lemma 2 is symmetrical to Lemma 1 for commitment creation and is used to prove Theorem 3 below.

Lemma 2. *Take a **L**-model M. Then, for every $w_{\sim x} \in W^{i:j:x-(\psi,\varphi)}$, we have: $M^{i:j:x-(\psi,\varphi)}, w_{\sim x} \models \chi$ iff $M, w \models \chi$.*

There are reduction axioms for commitment cancelation which are symmetrical to the reduction axioms for commitment creation.

Theorem 3. *The following equivalences are valid:*

T1. $[i{:}j{:}x{-}(\psi,\varphi)]\mathrm{wr}_{i,j,x} \leftrightarrow (\mathrm{wr}_{i,j,x} \wedge (\psi \rightarrow [i]\varphi))$
T2. $[i{:}j{:}x{-}(\psi,\varphi)]\alpha \leftrightarrow \alpha$ *if* $\alpha \neq \mathrm{wr}_{i,j,x}$
T3. $[i{:}j{:}x{-}(\psi,\varphi)]\neg\chi \leftrightarrow \neg[i{:}j{:}x{-}(\psi,\varphi)]\chi$
T4. $[i{:}j{:}x{-}(\psi,\varphi)](\chi_1 \wedge \chi_2) \leftrightarrow ([i{:}j{:}x{-}(\psi,\varphi)]\chi_1 \wedge [i{:}j{:}x{-}(\psi,\varphi)]\chi_2)$
T5. $[i{:}j{:}x{-}(\psi,\varphi)]\mathbf{D}_x\chi \leftrightarrow \mathbf{D}_x[i{:}j{:}x{-}(\psi,\varphi)]\chi$
T6. $[i{:}j{:}x{-}(\psi,\varphi)]\mathbf{D}_y\chi \leftrightarrow \mathbf{D}_y\chi$ *if* $y \neq x$
T7. $[i{:}j{:}x{-}(\psi,\varphi)][i]\chi \leftrightarrow [i][i{:}j{:}x{-}(\psi,\varphi)]\chi$
T8. $[i{:}j{:}x{-}(\psi,\varphi)]\Box\chi \leftrightarrow \Box[i{:}j{:}x{-}(\psi,\varphi)]\chi$

Some properties The following are examples of valid properties of commitment cancelation, with $i, j, h, k \in Agt$, $p, q \in Atm$, and $x, y \in Inst$:

(13) $[i{:}j{:}x{-}(q,p)]\neg\mathbf{C}_{i:j:x}(q,p)$
(14) $\neg\mathbf{C}_{h:k:x}(q,p) \rightarrow [i{:}j{:}x{-}(\psi,\varphi)]\neg\mathbf{C}_{h:k:x}(q,p)$
(15) $\mathbf{C}_{h:k:x}(q,p) \rightarrow [i{:}j{:}x{-}(\psi,\varphi)]\mathbf{C}_{h:k:x}(q,p)$ if $h \neq i$ or $k \neq j$
(16) $\mathbf{C}_{h:k:y}(\gamma,\chi) \rightarrow [i{:}j{:}x{-}(\psi,\varphi)]\mathbf{C}_{h:k:y}(\gamma,\chi)$ if $y \neq x$
(17) $\neg\mathbf{C}_{h:k:y}(\gamma,\chi) \rightarrow [i{:}j{:}x{-}(\psi,\varphi)]\neg\mathbf{C}_{h:k:y}(\gamma,\chi)$ if $y \neq x$

$\mathbf{L}^{dyn}$-theorem 13 highlights that the operators of commitment cancelation are also well-defined. Suppose p and q are propositional atoms in Atm. Then, after the cancelation from institution x of i's commitment towards j to ensure p if q holds, i will not be committed to j in x to ensure p if q holds. According to theorem 14, the operation of canceling a commitment from a given institution x does not create new commitments about propositions in the same institution. Note that this property is due to the fact that in the model resulting from the operation of commitment cancelation, the atom $\mathrm{wr}_{i,j,x}$ is set to true at a world w_x only if $\mathrm{wr}_{i,j,x}$ was already true at w in the original model. Consequently, the set of worlds in the 'x-copy' of the original model in which $\mathrm{wr}_{i,j,x}$ is true is smaller than the set of worlds in the original model in which $\mathrm{wr}_{i,j,x}$ is true. According to theorem 15, the operation of canceling a commitment about propositions from a debtor i to a creditor j in a given institution x does not cancel a pre-existent commitment in the same institution from a debtor h to a creditor k, if either h and i are different debtors or k and j are different creditors. Similarly to commitment creation, theorems 16 and 17 highlight some locality aspects of commitment cancelation, where locality means that the process of canceling a commitment from an institution x does not change the commitments in an institution different from x.

4.3 Completeness

I call $\mathbf{L}^{dyn}$ the logic axiomatized by the principles of the logic $\mathbf{L}$ *plus* the axiom schemata of Theorems 2 and 3 and the rule of replacement of proved equivalence. I write $\vdash_{\mathbf{L}^{dyn}} \varphi$ if φ is a $\mathbf{L}^{dyn}$-theorem. In order to prove that the logic $\mathbf{L}^{dyn}$ is complete we need first an expressiveness result.

Proposition 3. *For every $\boldsymbol{L}^{dyn}$-formula φ there exists an $\boldsymbol{L}$-formula φ' such that $\vdash_{\boldsymbol{L}^{dyn}} \varphi \leftrightarrow \varphi'$.*

Proof. By means of the principles **R1**-**R8** in Theorem 2 and **T1**-**T8** in Theorem 3, it is straightforward to prove that for every $\mathbf{L}^{dyn}$ formula there is an equivalent $\mathbf{L}$ formula. In fact, each reduction axiom **R3**-**R8** and **T3**-**T8**, when applied from the left to the right by means of the rule of replacement of proved equivalence, yields a simpler formula, where 'simpler' roughly speaking means that the dynamic operators are pushed inwards. Once the dynamic operators attain an atom they are eliminated by one of the equivalences **R1**-**R2** and **T1**-**T2**.

Theorem 4. *The logic $\boldsymbol{L}^{dyn}$ is complete.*

Proof. The theorem is a straightforward consequence of Theorem 1 and Proposition 3, together with the fact that the logic $\mathbf{L}^{dyn}$ is a conservative extension of the logic $\mathbf{L}$.

5 Related works and perspectives

The formal semantics for commitment creation and commitment cancelation proposed in Section 4 can be seen as an application of the logical theory of *assignments* [26, 24] that has been recently applied to model knowledge and intention dynamics (see, e.g., [25, 20]). It has also to be noted that the semantics of Section 4 corresponds to a two-event action model à la Baltag, Moss and Solecki (BMS) [3].

Several formal approaches to commitment have been recently proposed in the area of multi-agent systems and in the area of deontic logic. For instance, in [29] conditional commitments are modeled as unconditional commitments combined with strict (non material) implication. Bentahar et al. [6] provide an analysis of (propositional) dialectical and practical commitments in the context of conversation using a formal language based on an extended version of CTL (Computational Tree Logic) and on dynamic logic. A part of their work is devoted to discuss some intuitive reasoning postulates for commitment. Khan & Lésperance [18] propose a rich formal analysis of conditional commitment. They use the term 'intention' and 'commitment' interchangeably since, they argue, the logical structure of the two is similar enough. Their account is set within a framework for modeling communicating agents based on the Situation Calculus. Verdicchio & Colombetti [27] also formalize commitments in a variant of CTL. They formally define the concepts of fulfilled commitment, violated commitment and pending commitment. However, all these approaches do not provide a sound and complete logic for the analysis of the static and dynamics aspects of commitments.

A different approach is proposed by Singh [23] who takes seriously the issue of providing a model theory, and a sound and complete set of reasoning postulates for commitment. Singh focuses on dialectical (propositional) commitments and on practical commitments. However, its logical framework only applies to the static aspect of commitment and does not consider the dynamic dimension which is the main focus of the present contribution.

In this article I only considered unilateral commitments of an agent i towards another agent j. My main objective of future research is to extend the present analysis to

bilateral commitments in order to capture interesting social notions such as the notion of agreement and the notion of contract. In fact, as a first approximation, an agreement can be defined as a bilateral commitment among the agents in a group. In more formal terms, two agents i_1 and i_2 have an agreement in the context of institution x to ensure respectively φ_1 if ψ_1 holds and φ_2 if ψ_2 holds if and only if, i_1 is committed to i_2 to ensure φ_1 if ψ_1 holds, and i_2 is committed to i_1 to ensure φ_2 if ψ_2 holds, that is: $\mathbf{C}_{i_1:i_2:x}(\psi_1,\varphi_1) \wedge \mathbf{C}_{i_2:i_1:x}(\psi_2,\varphi_2)$. Another aspect I intend to investigate in the future is a generalization of the present framework to commitments of an agent towards a group, and to commitments of a group towards an agent (or towards a group).

I also postpone to future work an extension of the logic $\mathbf{L}$ and of its dynamic variant $\mathbf{L}^{dyn}$ by a temporal modal operator of the form $\mathbf{F}$, where $\mathbf{F}\varphi$ means "φ will be true at some point in the future". This extension will enable to redefine the commitment modality $\mathbf{C}_{i:j:x}\varphi$ introduced in Section 3 by the formula $\mathbf{D}_x(\neg\mathbf{F}[i]\varphi \rightarrow \mathrm{wr}_{i,j,x}) \wedge \widehat{\mathbf{D}}_x\neg\mathbf{F}[i]\varphi$ which makes explicit the temporal aspect of commitment. According to this definition, an agent i is committed to agent j in the context of institution x to ensure φ if and only if, i has a duty towards j in the context of institution x to ensure φ *at some point in the future*. This temporal variant of the notion of commitment is analogous to the notion of achievement goal proposed in the domain of BDI logics (see, e.g., [10]).

6 Acknowledgements

The author is grateful to the anonymous referees of DEON 2010, Tiago de Lima and Andreas Herzig for their helpful comments on the contents of this work.

References

1. A. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 22:100–103, 1958.
2. P. Balbiani, A. Herzig, and N. Troquard. Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.
3. A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. In *Proc. of TARK'98*, pages 43–56, 1998.
4. P. Bartha. Conditional obligation, deontic paradoxes, and the logic of agency. *Annals of Mathematics and Artificial Intelligence*, 9:1–23, 1993.
5. N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, New York, 2001.
6. J. Bentahar, B. Muolin, J.-J. Ch. Meyer, and B. Chaib-draa. A logical model for commitment and argument network for agent communication. In *Proc. of AAMAS 2004*, pages 792–799. ACM Press, 2004.
7. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
8. C. Castelfranchi. Commitment: from individual intentions to groups and organizations. In *Proc. of ICMAS'95*, pages 528–535. MIT Press, 1995.
9. B. J. Chellas. Time and modality in the logic of agency. *Studia Logica*, 51:485–517, 1992.
10. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
11. N. Desai, N. C. Narendra, and M. P. Singh. Checking correctness of business contracts via commitments. In *Proc. of AAMAS 2008*, pages 787–794. ACM Press, 2008.

12. F. Dignum. Autonomous agents with norms. *Artificial Intelligence and Law*, 7:69–79, 1999.
13. D. Grossi, J.-J. Ch Meyer, and F. Dignum. Classificatory aspects of counts-as: An analysis in modal logic. *Journal of Logic and Computation*, 16(5):613–643, 2006.
14. H. Herrestad and C. Krogh. Obligations directed from bearers to counterparties. In *Proc. of the Fifth International Conference on Artificial intelligence and law*, pages 210–218. ACM Press, 1995.
15. J. F. Horty and N. Belnap. The deliberative STIT: A study of action, omission, and obligation. *Journal of Philosophical Logic*, 24(6):583–644, 1995.
16. A. Jones and M. J. Sergot. A formal characterization institutionalised power. *Journal of the IGPL*, 4:429–445, 1996.
17. S. Kanger and H. Kanger. Rights and parliamentarism. *Theoria*, 6(2):85–115, 1966.
18. S. M. Khan and Y. Lésperance. On the semantics of conditional commitment. In *Proc. of AAMAS 2006*, pages 1337–1344, 2006.
19. L. Lindahl. Stig Kanger's theory of rights. In D. Prawitz, B. Skyrms, and D. Westeråhl, editors, *Logic, Methodology and Philosophy of Science*, volume IX. Elsevier, 1994.
20. E. Lorini, M. Dastani, H. van Ditmarsch, A. Herzig, and J.-J. Ch. Meyer. Intentions and assignments. In *Proc. of LORI 2009*, LNCS, pages 198–211, 2009.
21. D. Makinson. On the formal representation of rights relations: remarks on the work of Stig Kanger and Lars Lindahl. *The Journal of Philosophical Logic*, 15:403–425, 1986.
22. M. P. Singh. An ontology for commitments in multiagent systems. *Artificial Intelligence and Law*, 7:97–113, 1999.
23. M. P. Singh. Semantical considerations on dialectical and practical commitments. In *Proc. of AAAI'08*, pages 176–181. AAAI Press, 2008.
24. J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.
25. H. van Ditmarsch, A. Herzig, J. Lang, and P. Marquis. Introspective forgetting. *Synthese*, 169(2):405–423, 2009.
26. H. van Ditmarsch and B. Kooi. Semantic results for ontic and epistemic change. In *Proc. of LOFT 7*, pages 87–117, 2008.
27. M. Verdicchio and M. Colombetti. A logical model of social commitment for agent communication. In *Proc. of AAMAS 2003*, pages 528–535. ACM Press, 2003.
28. M. Xu. Axioms for deliberative STIT. *Journal of Philosophical Logic*, 27:505–552, 1998.
29. P. Yolum and M. Singh. Commitment machines. In *Proc. of ATAL'01*, volume 2333 of *LNAI*. Springer-Verlag, 2002.