Intention change via local assignments

Hans Van Ditmarsch, Tiago De Lima, Emiliano Lorini

Intention Change via Local Assignments*

Hans van Ditmarsch[1,**], Tiago de Lima[2], and Emiliano Lorini[3]

[1] University of Sevilla, Spain
[2] CRIL, University of Artois and CNRS, France
[3] IRIT, CNRS, France

Abstract. We present a logical approach to intention change. Inspired by Bratman's theory, we define intention as the choice to perform a given action at a certain time point in the future. This notion is modeled in a modal logic containing a temporal modality and modal operators for belief and choice. Intention change is then modeled by a specific kind of dynamic operator, that we call 'local assignment'. This is an operation on the model that changes the truth value of atomic formulae at specific time points. Two particular kinds of intention change are considered in some detail: intention generation and intention reconsideration.

1 Introduction

According to Bratman's planning theory of intention [6], rational agents build complex plans and organize their life on the basis of sequences of actions. They intend to perform certain actions and plans because they have reasons to perform them, and they write these actions and plans on their mental agenda, in order to remember when to perform them. In other terms, rational agents settle themselves in advance on plans for the future. That is, they have *future-directed intentions*. But also, they intend to do things here and now. That is, they have *present-directed intentions*. In such situations, they initiate the intended action and sustain it until its completion. As the time goes on, rational agents keep their future-oriented intentions unless they have no more reason to perform in the future what they intend to do. Hence, they reconsider their plans and possibly change them.

So, a future-directed intention has its own life in the mind of an agent. There is an initial moment in which it is generated. As time goes on, it may be reconsidered and eventually dropped. But, it may also last until it transforms into a present-directed intention, which is responsible for initiating the agent's action.

Since the seminal work of Cohen & Levesque [9] aimed at implementing Bratman's theory of intention, many formal logics for reasoning about intentions and plans, and for describing their dynamics have been developed (see, *e.g.* [23,27,22,29,20,2,7,10,18,15]). Most of them are based on dynamic logic extended by doxastic modal operators, and by modal operators for motivational attitudes, such as preferences, goals and intentions. These logics are traditionally called BDI (belief, desire, intention) logics. But, although

* This paper is an extended version of [21].

** Hans van Ditmarsch is also affiliated to the Institute of Mathematical Sciences Chennai (IMSC), as associated researcher.

logical analysis of intention and plan dynamics are available in the literature, the issue of a *formal semantics* for the dynamics of intentions and plans has received much less attention. Indeed, all previous approaches are mostly interested in characterizing in the object language the epistemic conditions under which an agent's intention persists over time, and the epistemic conditions under which an agent's intention is generated. However, they do not provide a semantic characterization of the process of generating an intention and of the process of reconsidering an intention.

The aim of this work is to shed light on this unexplored area by proposing a formal semantics of intention and plan dynamics based on the notion of *local assignment*. The function of a local assignment is to change the truth value of a given proposition at a specific time point along a history. We combine a static modal logic including a temporal modality and modal operators for mental attitudes belief and choice with three kinds of dynamic modalities and corresponding three kinds of local assignments: local assignments operating on an agent's beliefs, local assignments operating on the agent's choices and local assignments operating on the physical world. An agent's intention is defined in our approach as the agent's choice to perform a given action at a certain time point in the future, and two operations on intentions called *intention generation* and *intention reconsideration* are defined as specific kinds of local assignments on choices.

The rest of the paper is organized as follows. The first part (Section 2) introduces a static logic of time, action, belief, choice and intention. In the second part (Section 3), we move from a static perspective on agents' attitudes to a dynamic perspective, by adding the notion of local assignment to the logic of Section 2. We first present the syntax and semantics of three kinds of assignments: on beliefs, on choices and on the physical world. Then, in Section 4, we focus on two specific kinds of local assignment which allow to model the processes of plan generation and plan reconsideration. In Section 5, we apply our logical framework to a concrete example. Finally, in Section 6, we discuss some related work.

2 A logic of Time, Action and Mental Attitudes

We introduce a modal logic called **L** which supports reasoning about time, action and three different kinds of mental attitudes: beliefs, choices (or chosen goals), and intentions.

2.1 Syntax

Let $\mathbb{N} = \{0, 1, 2, \dots\}$ be the set of nonnegative integers. Let $ATM^{Fact} = \{f_1, f_2, \dots\}$ be a nonempty finite set of atoms denoting facts (or state of affairs). And let $ATM^{Act} = \{\alpha, \beta, \dots\}$ be a nonempty finite set of atoms denoting actions. The atom α stands for 'the agent performs a certain action α'. We define $ATM = ATM^{Fact} \cup ATM^{Act}$ to be the set of atomic formulae. We denote $p, q, \dots$ the elements in ATM.

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

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid [\mathsf{B}]\varphi \mid [\mathsf{C}]\varphi \mid \bigcirc\varphi$$

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

The three modal operators of our logic have the following reading: $[\mathtt{B}]\varphi$ means 'the agent believes that φ', $[\mathtt{C}]\varphi$ means 'the agent has chosen φ' (or 'the agent wants φ to be true'), and $\bigcirc\varphi$ means 'φ will be true in the next state, if no event affecting the world occurs'. The operator $\bigcirc$ describes the passive (or inertial) evolution of the world. That is, how the world evolves over time when no event affecting it occurs (this point will be better clarified in Section 3.4). [1] Operator $[\mathtt{C}]$ is used to denote the agent's choices. That is, the state of affairs that the agent has decided to pursue. Similar operators have been studied in [9,20,23]. We write $\bigcirc^n\varphi$ to indicate that the sentence φ is subject to n iterations of the modality $\bigcirc$, where $n \in \mathbb{N}$. More formally, $\bigcirc^0\varphi \overset{\text{def}}{=} \varphi$, and $\bigcirc^{n+1}\varphi \overset{\text{def}}{=} \bigcirc\bigcirc^n\varphi$. The following abbreviation defines the concept of intention for every $\alpha \in ATM^{Act}$ and $n \in \mathbb{N}$:

$$\mathtt{I}^n(\alpha) \overset{\text{def}}{=} [\mathtt{C}]\bigcirc^n\alpha$$

$\mathtt{I}^n(\alpha)$ stands for 'the agent intends to do action α in n steps from now'. Note that, if the agent has the intention to do α only once n steps from now, then after n steps $\mathtt{I}^n(\alpha)$ does not hold anymore. That is, the intention is dropped once n steps have passed.

2.2 Semantics

Definition 1 (L-model). *Models of the logic* $\boldsymbol{L}$ *(*$\boldsymbol{L}$*-models) are tuples* $M = \langle H, \mathscr{B}, \mathscr{C}, \mathscr{V}\rangle$*, where:*

- $H = \{h, h', \dots\}$ *is a nonempty set of possible histories;*
- $\mathscr{B}$ *and* $\mathscr{C}$ *are two total functions with signature* $H \longrightarrow 2^H$ *such that for every* $h \in H$*:*
  (C1) if $h' \in \mathscr{B}(h)$ *then* $\mathscr{B}(h') = \mathscr{B}(h)$*,*
  (C2) if $h' \in \mathscr{B}(h)$ *then* $\mathscr{C}(h') = \mathscr{C}(h)$*;*
- $\mathscr{V}$ *is a valuation function with signature* $ATM \longrightarrow 2^{H\times\mathbb{N}}$.

For every history h, $\mathscr{B}(h)$ is the set of histories that are compatible with the agent's beliefs at history h (or belief accessible histories at h), and $\mathscr{C}(h)$ is the set of histories that are compatible with the agent's choices at history h (or choice accessible histories at h). Constraint C1 (resp. C2) expresses that the agent's beliefs (resp. choices) are positively and negatively introspective.

Relations $\mathscr{C}$ and $\mathscr{B}$ are defined on histories instead of on points on histories. If the latter approach had been taken, it would be possible that the agent chooses, resp. believes, to be at a different time point than the actual one. To avoid this, it would be necessary to add restrictions to the model in order to "synchronize" choices and believes. The resulting class of models would validate the same formulae as the one we use here.

We call 'pointed model' a pair $M, h(n)$, where M is a model as defined above, $h \in H$ and $n \in \mathbb{N}$.

[1] The logic presented up till here could easily be extended with other temporal operators, such as "until". But, this would make the axiomatization of its extension, presented in Section 3.3, more complicated. We prefer to leave it for future work.

Definition 2 (Truth of L-formulae). *The satisfaction relation* $\models$*, between formulae in* **L** *and pointed models, is defined recursively as follows:*

$$\begin{array}{lll} M, h(n) \models \top & & \\ M, h(n) \models p & \textit{iff} & (h, n) \in \mathscr{V}(p) \\ M, h(n) \models \neg\varphi & \textit{iff} & \textit{not } M, h(n) \models \varphi \\ M, h(n) \models \varphi \vee \psi & \textit{iff} & M, h(n) \models \varphi \textit{ or } M, h(n) \models \psi \\ M, h(n) \models \bigcirc\varphi & \textit{iff} & M, h(n+1) \models \varphi \\ M, h(n) \models [\mathtt{C}]\varphi & \textit{iff} & M, h'(n) \models \varphi \textit{ for all } h' \in \mathscr{C}(h) \\ M, h(n) \models [\mathtt{B}]\varphi & \textit{iff} & M, h'(n) \models \varphi \textit{ for all } h' \in \mathscr{B}(h) \end{array}$$

We write $\models_{\mathbf{L}} \varphi$ to denote that φ is *valid* (*i.e.* φ is true in all **L**-pointed models). We say that φ is *satisfiable* if and only if $\neg\varphi$ is not valid.

2.3 Axiomatization

Fig. 1 (on page 140) contains the axiomatization of the logic **L**. We have Axioms K, D, 4 and 5 for beliefs and Axioms K and D for choices (as in [9]). Thus, we assume that if an agent believes (resp. does not believe) that φ then he believes this (Axioms 4 and 5 for $[\mathtt{B}]$), and we also assume that an agent cannot have inconsistent beliefs (Axiom D for $[\mathtt{B}]$). We assume that an agent cannot have inconsistent choices (Axiom D for $[\mathtt{C}]$). And we also assume that if an agent wants (resp. does not want) φ to be true then he believes this (Axioms $\mathbf{PIntr}_{[\mathtt{C}]}$ and $\mathbf{NIntr}_{[\mathtt{C}]}$). Similar principles of positive and negative introspection for choices are given in [13]. At the current stage, these are the only interaction principles between beliefs and choices. We postpone to future work a refinement of the logic **L** by interaction principles like $[\mathtt{B}]\varphi \rightarrow [\mathtt{C}]\varphi$ (if the agent believes that φ then he wants φ to be true) or $[\mathtt{C}]\varphi \rightarrow [\mathtt{B}]\varphi$ (if the agent wants φ to be true then he believes that φ). Similar principles have been studied for instance in [9,23].

We also have a basic principle for the temporal *next* operator (Axiom $\mathbf{Funct}_{\bigcirc}$): φ will be true in the next state if and only if it is not the case that φ will be false in the next state.

Finally, we have interaction principles between time and beliefs, and between time and goals. These are called *perfect recall* (Axioms $\mathbf{PR}_{[\mathtt{B}]}$ and $\mathbf{PR}_{[\mathtt{C}]}$) and *no learning* (Axioms $\mathbf{NL}_{[\mathtt{B}]}$ and $\mathbf{NL}_{[\mathtt{C}]}$) [14]. According to these four axioms, if the agent believes (resp. wants) that φ will be true in the next state if no event affecting the world occurs then, if no event affecting the world occurs, in the next state the agent will believe (resp. want) that φ and vice-versa.

We call **L** the logic axiomatized by the principles in Fig. 1, and we write $\vdash_{\mathbf{L}} \varphi$ if φ is a **L**-theorem. For example, the following is provable using Axiom 4, Necessitation rule for $[\mathtt{B}]$ and Axiom $\mathbf{PR}_{[\mathtt{B}]}$:

$$\vdash_{\mathbf{L}} [\mathtt{B}]\bigcirc\varphi \leftrightarrow [\mathtt{B}]\bigcirc[\mathtt{B}]\varphi$$

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

(**PC**)	All principles of classical propositional calculus
($\mathbf{KD45}_{[B]}$)	All principles of modal logic KD45 for $[B]$
($\mathbf{KD}_{[C]}$)	All principles of modal logic KD for $[C]$
($\mathbf{K}_{\bigcirc}$)	All principles of modal logic K for $\bigcirc$
($\mathbf{PIntr}_{[C]}$)	$[C]\varphi \rightarrow [B][C]\varphi$
($\mathbf{NIntr}_{[C]}$)	$\neg[C]\varphi \rightarrow [B]\neg[C]\varphi$
($\mathbf{Funct}_{\bigcirc}$)	$\bigcirc\varphi \leftrightarrow \neg\bigcirc\neg\varphi$
($\mathbf{PR}_{[B]}$)	$[B]\bigcirc\varphi \rightarrow \bigcirc[B]\varphi$
($\mathbf{PR}_{[C]}$)	$[C]\bigcirc\varphi \rightarrow \bigcirc[C]\varphi$
($\mathbf{NL}_{[B]}$)	$\bigcirc[B]\varphi \rightarrow [B]\bigcirc\varphi$
($\mathbf{NL}_{[C]}$)	$\bigcirc[C]\varphi \rightarrow [C]\bigcirc\varphi$

Fig. 1. Axiomatization of **L**

Proof. First, we provide an alternative semantics for **L**, in terms of standard Kripke frames. An alternative model is a tuple of the form $M' = \langle W, R_{\bigcirc}, R_{[B]}, R_{[C]}, V\rangle$, where W is a non-empty set of possible worlds, $R_{\bigcirc}$ is a serial and deterministic accessibility relation over W, $R_{[B]}$ is a serial, transitive and Euclidean accessibility relation over W, $R_{[C]}$ is a serial accessibility relation over W, V is a valuation function with signature $ATM \rightarrow 2^W$, and where the following interaction constraints are satisfied:

($\mathbf{PIntr}_{[B]}$)	if $wR_{[B]}w'$ and $w'R_{[C]}w''$ then $wR_{[C]}w''$;
($\mathbf{NIntr}_{[C]}$)	if $wR_{[B]}w'$ and $wR_{[C]}w''$ then $w'R_{[C]}w''$;
($\mathbf{PR}_{[B]}$)	if $w(R_{\bigcirc} \circ R_{[B]})w'$ then $w(R_{[B]} \circ R_{\bigcirc})w'$;
($\mathbf{PR}_{[C]}$)	if $w(R_{\bigcirc} \circ R_{[C]})w'$ then $w(R_{[C]} \circ R_{\bigcirc})w'$;
($\mathbf{NL}_{[B]}$)	if $w(R_{[B]} \circ R_{\bigcirc})w'$ then $w(R_{\bigcirc} \circ R_{[B]})w'$;
($\mathbf{NL}_{[C]}$)	if $w(R_{[C]} \circ R_{\bigcirc})w'$ then $w(R_{\bigcirc} \circ R_{[C]})w'$.

Alternative pointed models are tuples of the form $\langle M', w\rangle$ where M' is as defined above and $w \in W$. The alternative satisfaction relation $\models'$, as well as validity, are defined as usual.

Second, it is easy to see that the axiomatic system in Fig. 1 is sound and complete with respect to the class of alternative models, via the Sahlqvist theorem, cf. [5, Th. 2.42]. Indeed all axioms in Fig. 1 are in the so-called Sahlqvist class [25]. Thus, they are all expressible as first-order conditions on Kripke models and are complete with respect to the defined model classes.

Third, we show that for every alternative pointed model $\langle M', w\rangle$, there is a pointed model $\langle M, h\rangle$, where $M = \langle H, \mathscr{B}, \mathscr{C}, \mathscr{V}\rangle$, such that for every formula $\varphi \in \mathcal{L}$, $M', w \models' \varphi$ if and only if $M, h(0) \models \varphi$. The construction of $\langle M, h\rangle$ is performed in three steps:

Step 1: Unravel the alternative pointed model $\langle M', w\rangle$, in such a way that it becomes an infinite tree with root w and which is bisimilar [5] to the original model M'. Call this new tree-shaped model M''.

Step 2: Label each world v in M'' with a natural number $L(v)$, as follows: (a) $L(w) = 0$. (b) $L(v') = L(v)$, if $v' \in (R_{[\mathtt{B}]} \cup R_{[\mathtt{C}]})(v)$. (c) $L(v') = L(v) + 1$, if $v' \in R_{\bigcirc}(v)$. Remark 1. Because M' satisfy constraints $\mathbf{NL}_{[\mathtt{B}]}$ $\mathbf{PR}_{[\mathtt{C}]}$ (resp. $\mathbf{NL}_{[\mathtt{B}]}$ and $\mathbf{PR}_{[\mathtt{C}]}$), model M'' constructed in steps 1 and 2 has the following property: Let $(v, v'), (u, u') \in R^+_{\bigcirc}$ (the transitive closure of $R_{\bigcirc}$) such that $L(v) = L(u)$ and $L(v') = L(u')$. Then, $(v, u) \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$) if and only if $(v', u') \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$).

Step 3: Construct the model $M = \langle H, \mathscr{B}, \mathscr{C}, \mathscr{V} \rangle$ from model M'', as follows: (a) H is the set of branches $h = (v_0, v_1, v_2, \dots)$ of the tree such that $(v_i, v_{i+1}) \in R_{\bigcirc}$, for all $i \geq 0$. (b) $\mathscr{B}$ (resp. $\mathscr{C}$) is the set of pairs of branches (h, h') of H such that there is an arrow labeled by $R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$) from a world in h to a world in h', and (c) $\mathscr{V}(p)$ is the set of pairs (h, n) such that $h \in H$, v is in the branch h, $L(v) = n$ and $v \in V(p)$.

Remark 2. It follows from the construction of M'' that, if $(h, h') \in \mathscr{B}$ (resp. $\mathscr{C}$) then there is a pair of worlds (v, v') such that $L(v) = L(v')$ and $(v, v') \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$).

Fourth, we show that $\langle M'', w \rangle$ and $\langle M, h(0) \rangle$ are, in some sense, bisimilar.

Forth condition : It is easy to see that, by construction of M, if $(v, v') \in R_{\bigcirc}$ then $L(v') = L(v) + 1$. Moreover, let v be in branch $h \in H$, and v' in branch $h' \in H$. Then, if $(v, v') \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$) then, again by construction, $(h, h') \in \mathscr{B}$ (resp. $\mathscr{C}$).

Back condition : Analogously, it is easy to see that if $L(v') = L(v) + 1$ then $(v, v') \in R_{\bigcirc}$. Moreover, if $(h, h') \in \mathscr{B}$ (resp. $\mathscr{C}$) then, by Remark 1, there is a pair of worlds (v, v') in M'' such that $L(v) = L(v')$ and $(v, v') \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$). Moreover, by Remark 2, for every pair of worlds (u, u') such that u is in branch $h \in H$, u' is in branch $h' \in H$ and $L(u) = L(u')$, we have that $(u, u') \in R_{[\mathtt{B}]}$ (resp. $R_{[\mathtt{C}]}$).

Fifth, with an induction on the structure of φ we show that $\langle M'', v \rangle \models' \varphi$ if and only if $\langle M, h(L(v)) \rangle \models \varphi$, where v is in branch $h \in H$.

The induction base has two cases. Case 1 is $\varphi = \top$, and Case 2 is $\varphi = p$ for some atomic formula p. Both are straightforward. There are five cases in the induction step, one for each operator of the logic. The cases for the boolean operators $\neg$ and $\vee$ are straightforward.

Case 3 is $\varphi = \bigcirc\varphi_1$. $\langle M'', v \rangle \models' \bigcirc\varphi_1$ iff $\langle M'', v' \rangle \models' \varphi_1$, for all $v' \in R_{\bigcirc}(v)$ iff $\langle M, h(L(v') + 1) \rangle \models \varphi_1$ (by the forth and back conditions above and the induction hypothesis) iff $\langle M, h(L(v)) \rangle \models \bigcirc\varphi$.

Case 4 is $\varphi = [\mathtt{B}]\varphi_1$. $\langle M'', v \rangle \models' [\mathtt{B}]\varphi_1$ iff $\langle M'', v' \rangle \models' \varphi_1$, for all $v' \in R_{[\mathtt{B}]}(v)$ iff $\langle M, h'(L(v)) \rangle \models \varphi_1$, for all $h' \in R_{[\mathtt{B}]}(h)$, and where v' is in the sequence h'. (again, by the forth and back constructions above and the induction hypothesis) iff $\langle M, h(L(v)) \rangle \models [\mathtt{B}]\varphi_1$.

Case 5 is analogous to case Case 4.

Sixth, the converse can be done as well. That is, for every pointed model $\langle M, h \rangle$ it is possible to construct an alternative pointed model $\langle M', w \rangle$ that satisfies the same formulae of $\mathcal{L}$. This is straightforward, and left to the reader.

□

3 Local Assignments

In this section, we extend the logic **L** of Section 2 by modal operators for physical world change and mental attitude change. We distinguish two kinds of mental attitude change: belief change and choice change. We call $\mathbf{L}^+$ the extended logic. Logic $\mathbf{L}^+$ is based on the notion of *local assignment*. The function of a local assignment is to associate the truth value of a certain formula φ to a propositional atom p at a specific time point n along a history.

3.1 Syntax

We write *ASG* to denote the set of all partial functions σ with signature $(ATM \times \mathbb{N}) \to \mathcal{L}$. The elements in *ASG* are called *local assignments*, or simply *assignments*. We write *CASG* to denote the set of all triples $\Sigma = (\sigma_B, \sigma_C, \sigma_W)$ such that $\sigma_W, \sigma_B, \sigma_C \in ASG$. The elements in the set *CASG* are called *complex local assignments*, or simply *complex assignments*. Every complex assignment $\Sigma = (\sigma_B, \sigma_C, \sigma_W)$ is composed by a *belief assignment* σ_B (an assignment responsible for belief change), a *choice assignment* σ_C (an assignment responsible for choice change), and a *world assignment* σ_W (an assignment responsible for world change). When spelling out the elements of $\sigma_B = \{(p_1, n_1, \varphi_1), \ldots, (p_m, n_m, \varphi_m)\}$, we write it as $\{(p_1,n_1) \overset{B}{\mapsto} \varphi_1, \ldots, (p_m,n_m) \overset{B}{\mapsto} \varphi_m\}$, and analogously for σ_C and σ_W.

Definition 3 **($\Uparrow\Sigma$).** *Let $\Sigma = (\sigma_B,\sigma_C,\sigma_W)$. We define $\Uparrow\Sigma$ as the triple $(\Uparrow\sigma_B, \Uparrow\sigma_C, \Uparrow\sigma_W)$ such that, for all $p \in ATM$ and $n \in \mathbb{N}$:*

- $\Uparrow\sigma_B(p, n) = \sigma_B(p, n + 1)$,
- $\Uparrow\sigma_C(p, n) = \sigma_C(p, n + 1)$, *and*
- $\Uparrow\sigma_W(p, n) = \sigma_W(p, n + 1)$.

This means that the belief/choice/world assignments $\Uparrow\sigma_B/\Uparrow\sigma_C/\Uparrow\sigma_W$ are obtained by, respectively, shifting one step forward the belief/choice/world assignments $\sigma_B/\sigma_C/\sigma_W$.

The language $\mathcal{L}^+$ of the logic $\mathbf{L}^+$ is defined by the BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid [\mathsf{B}]\varphi \mid [\mathsf{C}]\varphi \mid \bigcirc\varphi \mid [\Sigma{:}W]\varphi \mid [\Sigma{:}B]\varphi \mid [\Sigma{:}C]\varphi$$

where p ranges over *ATM* and Σ ranges over *CASG*.

The formula $[\Sigma{:}W]\varphi$ stands for: 'φ holds in the physical world after the occurrence of the event Σ'. The formula $[\Sigma{:}B]\varphi$ stands for: 'φ holds in the context of the agent's beliefs after the occurrence of the event Σ'. (This does *not* mean that the agent believes φ after Σ. See Theorem 3 for precision.) The formula $[\Sigma{:}C]\varphi$ stands for: 'φ holds in the context of the agent's choices after the occurrence of the event Σ'.

3.2 Semantics

For every **L**-model M, every $n \in \mathbb{N}$ and every $\Sigma = (\sigma_B, \sigma_C, \sigma_W)$, we define the model M_n^Σ which results from the update of M at the time point n by the complex assignment Σ.

Definition 4 (Updated model M_n^Σ). *For every* $\mathbf{L}$*-model* $M = \langle H, \mathscr{B}, \mathscr{C}, \mathscr{V}\rangle$ *and every* $n \in \mathbb{N}$*,* M_n^Σ *is the tuple* $\langle H_n^\Sigma, \mathscr{B}_n^\Sigma, \mathscr{C}_n^\Sigma, \mathscr{V}_n^\Sigma\rangle$*, where:*

$$\begin{aligned}
H_n^\Sigma &= \{h_W | h \in H\} \cup \{h_B | h \in H\} \cup \{h_C | h \in H\};\\
\mathscr{B}_n^\Sigma(h_W) &= \{h'_B | h' \in \mathscr{B}(h)\};\\
\mathscr{B}_n^\Sigma(h_B) &= \{h'_B | h' \in \mathscr{B}(h)\};\\
\mathscr{B}_n^\Sigma(h_C) &= \{h'_C | h' \in \mathscr{B}(h)\};\\
\mathscr{C}_n^\Sigma(h_W) &= \{h'_C | h' \in \mathscr{C}(h)\};\\
\mathscr{C}_n^\Sigma(h_B) &= \{h'_C | h' \in \mathscr{C}(h)\};\\
\mathscr{C}_n^\Sigma(h_C) &= \{h'_C | h' \in \mathscr{C}(h)\};\\
\mathscr{V}_n^\Sigma(p) &= \{(h_W, k) | k \geq n \text{ and } M, h(k) \models \sigma_W(p, k-n)\} \cup\\
&\quad \{(h_W, k) | k < n \text{ and } M, h(k) \models p\} \cup\\
&\quad \{(h_B, k) | k \geq n \text{ and } M, h(k) \models \sigma_B(p, k-n)\} \cup\\
&\quad \{(h_B, k) | k < n \text{ and } M, h(k) \models p\} \cup\\
&\quad \{(h_C, k) | k \geq n \text{ and } M, h(k) \models \sigma_C(p, k-n)\} \cup\\
&\quad \{(h_C, k) | k < n \text{ and } M, h(k) \models p\}.
\end{aligned}$$

M_n^Σ is obtained by creating three copies of each history of the original model M (a copy for the physical world, a copy for belief, a copy for choice). Moreover, for every atom p and for every $k \in \mathbb{N}$ such that $k \geq n$, the effect of updating model M at the time point n by the event Σ is to assign the truth value of $\sigma_B(p, k-n)$ to the atom p at the time point k of all belief copies of the original histories, to assign the truth value of $\sigma_C(p, k-n)$ to the atom p at the time point k of all choice copies of the original histories, and to assign the truth value of $\sigma_W(p, k-n)$ to the atom p at the time point k of all world copies of the original histories. For example, suppose that $k = n$. Then, the effect of updating model M at the time point n by the event Σ, is to assign the truth value of $\sigma_B(p, 0)$ (resp. $\sigma_C(p, 0)$, resp. $\sigma_W(p, 0)$) to the atom p at the time point n of all belief copies (resp. choice copies, resp. world copies) of the original histories.

For every world copy h_W, at h_W the agent considers possible all belief copies of those histories that he considered possible before the event Σ, and he chooses all choice copies of those histories that he chose before the event Σ.

For every belief copy h_B, at h_B the agent considers possible all belief copies of those histories that he considered possible before the event Σ, and he chooses all choice copies of those histories that he chose before the event Σ.

For every choice copy h_C, at h_C the agent considers possible all choice copies of those histories that he considered possible before the event Σ, and he chooses all choice copies of those histories that he chose before the event Σ.

This construction of the updated model M_n^Σ ensures that the agent is aware that his choices have been changed accordingly so that the properties of positive and negative introspection over the agent's choices (Constraint C2 in Definition 1) are preserved after the occurrence of the event Σ.

Theorem 2. *If* M *is an* $\mathbf{L}$*-model then* M_n^Σ *is an* $\mathbf{L}$*-model.*

Proof. It is just trivial to prove that our operation of model update preserves Constraint C1 in Definition 1. Let us prove that it also preserves Constraint C2.

Assume $h'_B \in \mathscr{B}_n^\Sigma(h_W)$ and $h''_C \in \mathscr{C}_n^\Sigma(h'_B)$. It follows that $h' \in \mathscr{B}(h)$ and $h'' \in \mathscr{C}(h')$. Then, by constraint C2, we have $h'' \in \mathscr{C}(h)$. Therefore, $h''_C \in \mathscr{C}_n^\Sigma(h_W)$. In a similar way we can prove that if $h'_B \in \mathscr{B}_n^\Sigma(h_B)$ and $h''_C \in \mathscr{C}_n^\Sigma(h'_B)$ then $h''_C \in \mathscr{C}_n^\Sigma(h_B)$, and if $h'_C \in \mathscr{B}_n^\Sigma(h_C)$ and $h''_C \in \mathscr{C}_n^\Sigma(h'_C)$ then $h''_C \in \mathscr{C}_n^\Sigma(h_C)$.

Now, assume $h'_C \in \mathscr{C}_n^\Sigma(h_W)$ and $h''_B \in \mathscr{B}_n^\Sigma(h_W)$. It follows that $h' \in \mathscr{C}(h)$ and $h'' \in \mathscr{B}(h)$. Then, by constraint C2, we have $h' \in \mathscr{C}(h'')$. Therefore, $h'_C \in \mathscr{C}_n^\Sigma(h''_B)$. In a similar way we can prove that if $h'_C \in \mathscr{C}_n^\Sigma(h_B)$ and $h''_B \in \mathscr{B}_n^\Sigma(h_B)$ then $h'_C \in \mathscr{C}_n^\Sigma(h''_B)$, and if $h'_C \in \mathscr{C}_n^\Sigma(h_C)$ and $h''_C \in \mathscr{B}_n^\Sigma(h_C)$ then $h'_C \in \mathscr{C}_n^\Sigma(h''_C)$. □

Definition 5 (Truth of L$^+$-formulae). *The satisfaction relation* $\models$, *between formulae in* $\boldsymbol{L}^+$ *and pointed models, is defined by the conditions in Definition 2 together with the following three conditions:*

- $M, h(n) \models [\Sigma{:}\textsc{w}]\varphi$ *iff* $M_n^\Sigma, h_W(n{+}1) \models \varphi$;
- $M, h(n) \models [\Sigma{:}\textsc{b}]\varphi$ *iff* $M_n^\Sigma, h_B(n{+}1) \models \varphi$;
- $M, h(n) \models [\Sigma{:}\textsc{c}]\varphi$ *iff* $M_n^\Sigma, h_C(n{+}1) \models \varphi$.

Note that, according to the previous definition, the occurrence of the event Σ takes time. That is, an event Σ is a transition from a time point n along a history h of a model M to the *successor* of time point n along the world copy (or belief copy, or choice copy) of history h in the updated model M_n^Σ.

3.3 Axiomatization

We have reduction axioms for the three operators $[\Sigma{:}\textsc{w}]$, $[\Sigma{:}\textsc{b}]$ and $[\Sigma{:}\textsc{c}]$. They are called reduction axioms because, read from left to right, they reduce the complexity of those operators in a formula.

Theorem 3. *Suppose* $\Sigma = (\sigma_B, \sigma_C, \sigma_W)$. *Then, the following schemata are valid in* $\boldsymbol{L}^+$*:*

R1a.	$[\Sigma{:}\textsc{w}]p \leftrightarrow \sigma_W(p,1)$	***R4a.***	$[\Sigma{:}\textsc{w}]\bigcirc\varphi \leftrightarrow \bigcirc[\Uparrow\Sigma{:}\textsc{w}]\varphi$
R1b.	$[\Sigma{:}\textsc{b}]p \leftrightarrow \sigma_B(p,1)$	***R4b.***	$[\Sigma{:}\textsc{b}]\bigcirc\varphi \leftrightarrow \bigcirc[\Uparrow\Sigma{:}\textsc{b}]\varphi$
R1c.	$[\Sigma{:}\textsc{c}]p \leftrightarrow \sigma_C(p,1)$	***R4c.***	$[\Sigma{:}\textsc{c}]\bigcirc\varphi \leftrightarrow \bigcirc[\Uparrow\Sigma{:}\textsc{c}]\varphi$
R2a.	$[\Sigma{:}\textsc{w}]\neg\varphi \leftrightarrow \neg[\Sigma{:}\textsc{w}]\varphi$	***R5a.***	$[\Sigma{:}\textsc{w}][\mathtt{B}]\varphi \leftrightarrow [\mathtt{B}][\Sigma{:}\textsc{b}]\varphi$
R2b.	$[\Sigma{:}\textsc{b}]\neg\varphi \leftrightarrow \neg[\Sigma{:}\textsc{b}]\varphi$	***R5b.***	$[\Sigma{:}\textsc{b}][\mathtt{B}]\varphi \leftrightarrow [\mathtt{B}][\Sigma{:}\textsc{b}]\varphi$
R2c.	$[\Sigma{:}\textsc{c}]\neg\varphi \leftrightarrow \neg[\Sigma{:}\textsc{c}]\varphi$	***R5c.***	$[\Sigma{:}\textsc{c}][\mathtt{B}]\varphi \leftrightarrow [\mathtt{B}][\Sigma{:}\textsc{c}]\varphi$
R3a.	$[\Sigma{:}\textsc{w}](\varphi \wedge \psi) \leftrightarrow ([\Sigma{:}\textsc{w}]\varphi \wedge [\Sigma{:}\textsc{w}]\psi)$	***R6a.***	$[\Sigma{:}\textsc{w}][\mathtt{C}]\varphi \leftrightarrow [\mathtt{C}][\Sigma{:}\textsc{c}]\varphi$
R3b.	$[\Sigma{:}\textsc{b}](\varphi \wedge \psi) \leftrightarrow ([\Sigma{:}\textsc{b}]\varphi \wedge [\Sigma{:}\textsc{b}]\psi)$	***R6b.***	$[\Sigma{:}\textsc{b}][\mathtt{C}]\varphi \leftrightarrow [\mathtt{C}][\Sigma{:}\textsc{c}]\varphi$
R3c.	$[\Sigma{:}\textsc{c}](\varphi \wedge \psi) \leftrightarrow ([\Sigma{:}\textsc{c}]\varphi \wedge [\Sigma{:}\textsc{c}]\psi)$	***R6c.***	$[\Sigma{:}\textsc{c}][\mathtt{C}]\varphi \leftrightarrow [\mathtt{C}][\Sigma{:}\textsc{c}]\varphi$

Proof. We prove only **R4a** as an example.
$M, h(n) \models [\Sigma{:}\textsc{w}]\bigcirc\varphi$
IFF $M_n^\Sigma, h_W(n{+}1) \models \bigcirc\varphi$
IFF $M_n^\Sigma, h_W((n{+}1){+}1) \models \varphi$
IFF $M_{n+1}^{\Uparrow\Sigma}, h_W((n{+}1){+}1) \models \varphi$ (because $M_n^\Sigma = M_{n+1}^{\Uparrow\Sigma}$)
IFF $M, h(n{+}1) \models [\Uparrow\Sigma{:}\textsc{w}]\varphi$
IFF $M, h(n) \models \bigcirc[\Uparrow\Sigma{:}\textsc{w}]\varphi$. □

Theorem 4. *The logic* $\mathbf{L}^+$ *is completely axiomatized by principles in Fig. 1 together with the schemata of Theorem 3 and the rule of replacement of proved equivalence.*

Proof. Using the reduction axioms **R1a-R6c** in Theorem 3, and the rule of replacement of proved equivalence, every $\mathbf{L}^+$ formula can be reduced to an equivalent $\mathbf{L}$ formula. Hence, the completeness of $\mathbf{L}^+$ is a straightforward consequence of Theorem 1. □

In the rest of the paper we write $\vdash_{\mathbf{L}^+} \varphi$ if φ is a $\mathbf{L}^+$-theorem.

3.4 Discussion

The present approach offers two different temporal perspectives on world evolution, where the world includes both the physical world and the mental world (*i.e.* the agent's beliefs and choices).

In order to describe the *passive (or inertial) evolution of the world* (*i.e.* how the world evolves over time when there are no occurrences of events which affect the physical world and the agent's beliefs and choices), we use the *next* operator $\bigcirc$ in the static framework $\mathbf{L}$. Note that the operator $\bigcirc$ corresponds in the semantics to a transition from a state in the current model M to the *unique* successor state in the same model M.

In order to describe the *active evolution of the world* (*i.e.* how the world evolves over time when there are occurrences of events which affect the physical world and the agent's beliefs and choices), we use the dynamic operators $[\Sigma{:}w]$, $[\Sigma{:}B]$ and $[\Sigma{:}C]$ in the dynamic framework $\mathbf{L}^+$. We consider all possible transitions from the current model M to a new model which corresponds to the update of M through an event Σ affecting the physical world and the agent's beliefs and choices.

In other terms, in order to describe the *active evolution of the world* a branching time perspective is adopted in our approach. That is, we suppose that the world might *actively* evolve in many different ways depending on the event Σ which occurs and which affects it. On the contrary, in order to describe the *inertial evolution of the world* a linear time perspective is adopted in our approach, that is, we suppose that the world *passively/inertially* evolves in a deterministic way.

4 Intention and Plan Dynamics in $\mathbf{L}^+$

Two basic operations on an agent's intentions can be defined in $\mathbf{L}^+$: the operation of *generating* an intention to do an action α n steps from now, noted $gen(\alpha,n)$; and the operation of *reconsidering* (or *erasing*) an intention to do an action α n steps from now, noted $rec(\alpha,n)$. These two operations are defined as follows by means of choice assignments: [2]

$$gen(\alpha,n) \stackrel{\text{def}}{=} (\alpha,n) \stackrel{C}{\mapsto} \top$$
$$rec(\alpha,n) \stackrel{\text{def}}{=} (\alpha,n) \stackrel{C}{\mapsto} \bot$$

[2] The generalization to conjunctive intentions can be done with simultaneous assignments. For instance, operation $gen(\alpha \wedge \beta,n) \stackrel{\text{def}}{=} (\alpha,n) \stackrel{C}{\mapsto} \top, (\beta,n) \stackrel{C}{\mapsto} \top$, *i.e.* a partial local assignment function with domain $\{(\alpha,n),(\beta,n)\}$.

The following are $\mathbf{L}^{+}$-theorems which highlight some interesting properties of intention generation and intention reconsideration. For every $n, m \in \mathbb{N}$ and for every $\alpha, \beta \in ACT$ we have:

(1) $\vdash_{\mathbf{L}^{+}} [(\emptyset,\{gen(\alpha,n{+}1)\},\emptyset){:}w]\mathtt{I}^{n}(\alpha)$

(2) $\vdash_{\mathbf{L}^{+}} [(\emptyset,\{rec(\alpha,n{+}1)\},\emptyset){:}w]\neg\mathtt{I}^{n}(\alpha)$

(3) $\vdash_{\mathbf{L}^{+}} [(\emptyset,\{gen(\alpha,n{+}1)\},\emptyset){:}c]\bigcirc^{n}\alpha$

(4) $\vdash_{\mathbf{L}^{+}} [(\emptyset,\{rec(\alpha,n{+}1)\},\emptyset){:}c]\neg\bigcirc^{n}\alpha$

(5) $\vdash_{\mathbf{L}^{+}} \neg\mathtt{I}^{m}(\beta) \rightarrow [(\emptyset,\{gen(\alpha,n)\},\emptyset){:}w]\neg\mathtt{I}^{m-1}(\beta)$
if $\alpha \neq \beta$ or $m \neq n$

(6) $\vdash_{\mathbf{L}^{+}} \mathtt{I}^{m}(\beta) \rightarrow [(\emptyset,\{rec(\alpha,n)\},\emptyset){:}w]\mathtt{I}^{m-1}(\beta)$
if $\alpha \neq \beta$ or $m \neq n$

Proof. We prove theorem (1) as an example. Formula $[(\emptyset,\{gen(\alpha,n{+}1)\},\emptyset){:}w]\mathtt{I}^{n}(\alpha)$ is equivalent to $[\mathtt{C}][(\emptyset,\{gen(\alpha,n{+}1)\},\emptyset){:}c]\bigcirc^{n}\alpha$ (by **R6a** and rule of replacement of proved equivalences). The latter is equivalent to $[\mathtt{C}]\bigcirc^{n}[(\emptyset,\{gen(\alpha,1)\},\emptyset){:}c]\alpha$ (by repeated application of **R4c** and rule of replacement of proved equivalence) which in turn is equivalent to $[\mathtt{C}]\bigcirc^{n}\top$ (by **R1c** and rule of replacement of proved equivalence). The latter is equivalent to $\top$. □

According to theorem (1), after generating the intention to do α $n{+}1$ steps from now, in the physical world the agent intends to do α n steps from now. According to theorem (2), after reconsidering the intention to do α $n{+}1$ steps from now, in the physical world the agent does not intend to do α n steps from now. In Definition 5 we have supposed that the occurrence of a local assignment takes time (one time unit). Consequently, also the processes of generating/reconsidering an intention takes time. This is the reason why, as stated by theorems (1) and (2), the process of generating/reconsidering the intention to do α $n{+}1$ steps from now generates/reconsiders an intention to α n steps from now, and not an intention to do α $n{+}1$ steps from now.

Note that the two processes of intention generation and intention reconsideration comply with temporal precedence, that is, the process of reconsidering a certain intention cancels the effects of a previous process of generating the same intention, and the process of generating a certain intention cancels the effects of a previous process of reconsidering the same intention. More formally, by theorems (1) and (2), we have:

$$\vdash_{\mathbf{L}^{+}} [(\emptyset,\{gen(\alpha,n{+}2)\},\emptyset){:}w][(\emptyset,\{rec(\alpha,n{+}1)\},\emptyset){:}w]\neg\mathtt{I}^{n}(\alpha)$$
$$\vdash_{\mathbf{L}^{+}} [(\emptyset,\{rec(\alpha,n{+}2)\},\emptyset){:}w][(\emptyset,\{gen(\alpha,n{+}1)\},\emptyset){:}w]\mathtt{I}^{n}(\alpha)$$

Theorems (3) and (4) express the corresponding effects of the processes of intention generation and of intention reconsideration in the context of the agent's choices: after generating (resp. reconsidering) the intention to do α $n{+}1$ steps from now, in the context of the agent's choices it is the case that the agent will perform (resp. will not perform) action α n steps from now.

Theorems (5) and (6) express that the operations of intention generation and of intention reconsideration are characterized by partial modifications of an agent's plan. That is, the process of generating/reconsidering a plan does not affect the other plans of the agent: if α and β are different actions or m and n are different, and the agent intends (resp. does not intend) to do β m steps from now then, after reconsidering (resp. generating) the intention do α n steps from now, the agent will intend (resp. not intend) to do β $m-1$ steps from now.

Intention generation ($gen(\alpha,n)$) and reconsideration ($rec(\alpha,n)$) are mental events which have to be distinguished from the processes of starting an action (or trying to do an action) and stopping an action which are events operating on the physical world. The latter are defined by means of world assignments as follows:

$$start(\alpha) \stackrel{\text{def}}{=} (\alpha,1) \stackrel{W}{\mapsto} \top$$
$$stop(\alpha) \stackrel{\text{def}}{=} (\alpha,1) \stackrel{W}{\mapsto} \bot$$

The following are two $\mathbf{L}^+$-theorems which highlight the basic properties of the processes of starting an action and stopping an action. For every $\alpha \in ACT$ we have:

(7) $\vdash_{\mathbf{L}^+} [(\emptyset,\emptyset,\{start(\alpha)\}):w]\alpha$

(8) $\vdash_{\mathbf{L}^+} [(\emptyset,\emptyset,\{stop(\alpha)\}):w]\neg\alpha$

According to theorem (7), after starting action α, the agent performs action α. According to theorem (8), after stopping action α, the agent does not perform action α.

5 Application

Before concluding, we illustrate through an example how $\mathbf{L}^+$ can be concretely used to model intention dynamics.

Executability of intention generation. We denote with $\langle\langle Gen(\alpha,n)\rangle\rangle\varphi$ the fact 'it is possible that the agent will generate the intention to do action α n steps from now, and φ will be true afterwards'. Consequently, $\langle\langle Gen(\alpha,n)\rangle\rangle\top$ just means 'the agent will possibly generate the intention to do action α n steps from now'. The construction $\langle\langle Gen(\alpha,n)\rangle\rangle\varphi$ is defined as follows.

$$\langle\langle Gen(\alpha,n)\rangle\rangle\varphi \stackrel{\text{def}}{=} \neg \mathtt{I}^n(\alpha) \wedge [\mathtt{B}]\bigcirc^n good_\alpha \wedge [(\emptyset,\{gen(\alpha,n)\},\emptyset):w]\varphi$$

According to this definition, the *executability* of the process of generating the intention to do α n steps from now is determined by two conditions: the agent does not have already this intention (*i.e.* $\neg\mathtt{I}^n(\alpha)$) and he believes that, n steps from now, doing action α will be something good for him (*i.e.* $[\mathtt{B}]\bigcirc^n good_\alpha$), where $good_\alpha$ is a special atom in ATM^{Fact} expressing that 'performing action α is good for the agent'. Indeed, $\langle\langle Gen(\alpha,n)\rangle\rangle\top$ and $\neg\mathtt{I}^n(\alpha) \wedge [\mathtt{B}]\bigcirc^n good_\alpha$ are logically equivalent. Note that the condition $[\mathtt{B}]\bigcirc^n good_\alpha$ corresponds to the notion of *reasons for intending* or *reasons for acting*. This notion has been extensively studied in the philosophical literature on action and intention (see, *e.g.* [28,20]). A reason for intending is a belief that the agent uses as premise of a practical argument (*viz.* the argument that concludes in an intention).

Executability of intention reconsideration. The notion of executability of the process of intention reconsideration is defined in a similar way as follows:

$$\langle\langle Rec(\alpha,n)\rangle\rangle\varphi \stackrel{\mathtt{def}}{=} \mathtt{I}^n(\alpha) \wedge [\mathtt{B}]\bigcirc^n\neg good_\alpha \wedge \\ [(\emptyset,\{rec(\alpha,n)\},\emptyset){:}w]\varphi$$

$\langle\langle Rec(\alpha,n)\rangle\rangle\varphi$ denotes the fact 'it is possible that the agent will reconsider his intention to do action α n steps from now, and φ will be true afterwards'. Consequently, $\langle\langle Gen(\alpha,n)\rangle\rangle\top$ just means 'the agent will possibly reconsider his intention to do action α n steps from now'. According to this definition, the *executability* of the process of reconsidering the intention to do α n steps from now is determined by two conditions: (1) the agent has this intention (*i.e.* $\mathtt{I}^n(\alpha)$) and (2) he believes that, n steps from now, doing action α will be something bad for him (*i.e.* $[\mathtt{B}]\bigcirc^n\neg good_\alpha$). [3] Indeed, $\langle\langle Rec(\alpha,n)\rangle\rangle\top$ and $\mathtt{I}^n(\alpha) \wedge [\mathtt{B}]\bigcirc^n\neg good_\alpha$ are logically equivalent.

Executability of action. The notion of executability of the process of starting an action is defined as follows:

$$\langle\langle Start(\alpha)\rangle\rangle\varphi \stackrel{\mathtt{def}}{=} \mathtt{I}^1(\alpha) \wedge [\mathtt{B}]\bigcirc good_\alpha \wedge \\ [(\emptyset,\emptyset,\{start(\alpha)\}){:}w]\varphi$$

$\langle\langle Start(\alpha)\rangle\rangle\varphi$ denotes the fact 'it is possible that the agent will start action α, and φ will be true afterwards'. Consequently, $\langle\langle Gen(\alpha,n)\rangle\rangle\top$ just means 'the agent will possibly start action α'. According to this definition, the executability of the process of starting action α is determined by two conditions: (1) the agent has intention to perform action α in the next step ($\mathtt{I}^1(\alpha)$) and (2) he believes that, in the next state, doing action α will be something good for him (*i.e.* $[\mathtt{B}]\bigcirc good_\alpha$). Indeed, $\langle\langle Start(\alpha)\rangle\rangle\top$ and $\mathtt{I}^1(\alpha)\wedge[\mathtt{B}]\bigcirc good_\alpha$ are logically equivalent.

An example. We suppose that the agent wants to reach a certain place called *Utopia* in n steps from now, with $n \geq 4$. He can go to *Utopia* either by train or by car. That is, $ATM^{Act} = \{train, car\}$. The agent has decided to go by train. Thus, he has the intention to go to *Utopia* by train and does not have the intention to go to *Utopia* by car:

H. $\mathtt{I}^n(train) \wedge \neg\mathtt{I}^n(car)$.

Now, suppose the agent is informed that there is a train strike that day and, n steps from now, there will be no train going to *Utopia*. Thus, the agent learns that going to *Utopia* by train is a bad solution, whereas going to *Utopia* by car is a good solution. How are the agent's intentions and plans affected by this new information? The following theorem clarifies this point.

(9) $$\vdash_{\mathbf{L}^+} \mathbf{H} \rightarrow [(\{(good_{train},n) \stackrel{B}{\mapsto} \bot, (good_{car},n) \stackrel{B}{\mapsto} \top\},\emptyset,\emptyset){:}w] \\ \langle\langle Rec(train,n{-}1)\rangle\rangle\langle\langle Gen(car,n{-}2)\rangle\rangle \\ \bigcirc^{n-4}\langle\langle Start(car)\rangle\rangle car$$

[3] We here suppose that the agent believes that doing α will be something *bad* for him if and only if, he believes that doing α will be something *not good* for him.

According to theorem (9), hypothesis **H** ensures that, after having learnt that going to *Utopia* by train is a bad solution and going to *Utopia* by car is a good solution:

– the agent will possibly reconsider its intention to go to *Utopia* by train $n-1$ steps from now and,
– after that, it will possibly generate the intention to go to *Utopia* by car $n-2$ steps from now and,
– $n-4$ steps later, if the world will evolve passively/inertially, the agent will possibly start to go to *Utopia* by car and,
– as a consequence, he will go to *Utopia* by car.

Thus, theorem (9) shows that $\mathbf{L}^+$ concretely models intention dynamics in this example.

6 Related Work and Perspectives

Assignments were studied before in the literature on logic for information dynamics. However they were only applied to the dynamics of belief and knowledge [12,3], and there is still no application of this notion to the theory of intention. Furthermore, previous works in the area of Dynamic Epistemic Logic (DEL) [11] have only focused on assignments at a specific moment in time, and have not considered delayed assignments, that can operate locally, *i.e.* on specific future points in a model. In this paper we have shown that local assignments are well-suited to model intention and plan dynamics.

It has also to be noted that complex assignments restricted to $n = 0$ correspond to a three-event BMS action model [1]. Given a future point n in time, one can in principle again construct an action model from the composition of complex assignments and next operators in a sequence of length n (where the next operators correspond to 'clock ticks', so-called 'nothing happens' action models).

In [4,19] a logic of knowledge and preference dynamics is provided. In van Benthem & Liu's approach knowledge dynamics are modeled by means of announcements (or updates), whereas preference dynamics are modeled by means of operations on accessibility relations called upgrades. We think that local assignments, rather than announcements and upgrades, are more suited to model intention dynamics. Indeed, intention dynamics are obtained by partial modifications of an agent's plan and not necessarily by global modifications, and local assignments are a natural candidate to formalize these kinds of operations. This aspect of intention dynamics has been discussed in Section 3.3.

In [16], van der Hoek *et al.* present a formal model of intention revision strongly inspired by [9]. In this model an agent's mental state (which includes the agent's beliefs, desires and intentions) changes because the agent has made observations of his environment. In particular, observations cause change in beliefs and, indirectly, may produce change in the agent's intentions. The treatment of beliefs, desires, intentions and their dynamics proposed by van der Hoek *et al.* is rather syntactical (*i.e.* an agent's beliefs, desires and intentions are just sets of sentences and there are no modalities for these mental attitudes interpreted by means of model structures). This is the main difference with our approach whose objective is to provide a model-theoretic semantics of intentions and intention dynamics.

In [26], Shoham discusses the interaction of belief revision and intention revision. He treats belief revision more in the traditional AGM sense than in the dynamic epistemic modal sense. The paper has the character of a requirements analysis. A prior paper on belief and intention interaction in the AGM tradition, however without intention dynamics, is [8]. Work by Icard, Pacuit and Shoham [17] employs a dynamic modal approach for intention revision.

Rodenhäuser [24] associates an intention with a protocol, a sequence of planned actions that may also involve factual change—but not the local change as in our case; he also does not employ choice modalities.

Directions for future research are manifold. In this paper we only considered the single-agent case. We plan to extend our approach to the multi-agent setting in which agents can act in parallel and communicate their choices, beliefs and intentions to other agents. As anticipated in Section 2.3, we also plan to refine our logical framework by adding more interaction principles between the belief operator and the choice operator.

References

1. Baltag, A., Moss, L., Solecki, S.: The logic of public announcements, common knowledge and private suspicions. In: Proceedings of TARK 1998, pp. 43–56. Morgan Kaufmann, San Francisco (1998)
2. Bell, J., Huang, Z.: Dynamic goal hierarchies. Intelligent Agent Systems Theoretical and Practical Issues, 88–103 (1997)
3. van Benthem, J., van Eijck, J., Kooi, B.: Logics of communication and change. Information and Computation 204(11), 1620–1662 (2006)
4. van Benthem, J., Liu, F.: Dynamic logic of preference upgrade. Journal of Applied Non-Classical Logics 17(2), 157–182 (2007)
5. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001)
6. Bratman, M.: Intentions, plans, and practical reason. Harvard University Press, Cambridge (1987)
7. Broersen, J., Dastani, M., Hulstijn, J., van der Torre, L.: Goal generation in the BOID architecture. Cognitive Science Quarterly 2(3-4), 431–450 (2002); special issue on Desires, goals, intentions, and values: Computational architectures
8. Cleaver, T.W., Sattar, A.: Intention guided belief revision. In: Proceedings of AAAI 2007, pp. 36–41. AAAI Press, Menlo Park (2007)
9. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial Intelligence 42, 213–261 (1990)
10. Dignum, F., Kinny, D.: From desires, obligations and norms to goals. Cognitive Science Quarterly 2(3-4), 407–430 (2002)
11. van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Synthese Library Series, vol. 337. Springer, Heidelberg (2007)
12. van Ditmarsch, H., Kooi, B.: Semantic results for ontic and epistemic change. In: Logic and the Foundations of Game and Decision Theory (LOFT 7)., pp. 87–117. Texts in Logic and Games, Amsterdam University Press (2008)
13. Dunin-Keplicz, B., Verbrugge, R.: Collective intentions. Fundamenta Informaticae 51(3), 271–295 (2002)
14. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. MIT Press, Cambridge (1995)

15. Georgeff, M.P., Rao, A.S.: The semantics of intention maintenance for rational agents. In: Proceedings of IJCAI 1995, pp. 704–710 (1995)
16. Van der Hoek, W., Jamroga, W., Wooldridge, M.: Towards a theory of intention revision. Synthese 155(2), 265–290 (2007)
17. Icard, T., Pacuit, E., Shoham, Y.: Joint revision of beliefs and intention. In: Proceedings of KR 2010, pp. 572–574 (2010)
18. Khan, S.M., Lesperance, Y.: A logical framework for prioritized goal change. In: Proceedings of AAMAS 2010, pp. 283–290 (2010)
19. Liu, F.: Changing for the Better: Preference Dynamics and Agent Diversity. Ph.D. thesis, University of Amsterdam (2008)
20. Lorini, E., Herzig, A.: A logic of intention and attempt. Synthese 163(1), 45–77 (2008)
21. Lorini, E., van Ditmarsch, H., de Lima, T.: Logical model of intention and plan dynamics. In: Coelho, H., Studer, R., Wooldridge, M. (eds.) Proceedings of ECAI 2010, pp. 1075–1076. IOS Press, Amsterdam (2010)
22. Meyer, J.J.C., van der Hoek, W., van Linder, B.: A logical approach to the dynamics of commitments. Artificial Intelligence 113(1-2), 1–40 (1999)
23. Rao, A.S., Georgeff, M.P.: Modelling rational agents within a BDI-architecture. In: Proceedings of KR 1991, pp. 473–484. Morgan Kaufmann, San Francisco (1991)
24. Rodenhäuser, B.: Intentions in interaction. In: Proceedings of LOFT 2010 (2010)
25. Sahlqvist, H.: Completeness and correspondence in the first and second order semantics for modal logics. In: Proceedings of the 3rd Scandinavian Logic Symposium 1973, Studies in Logic, vol. (82) (1975)
26. Shoham, Y.: Logics of intention and the database perspective. Journal of Philosophical Logic 38(6), 633–647 (2009)
27. Singh, M., Asher, N.: A logic of intentions and beliefs. Journal of Philosophical Logic 22, 513–544 (1993)
28. Von Wright, G.H.: On so-called practical inference. The Philosophical Review 15, 39–53 (1972)
29. Wooldridge, M.: Reasoning about rational agents. MIT Press, Cambridge (2000)