A Logical Model of Intention and Plan Dynamics

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

Abstract. We propose 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 operating on an agent's beliefs, on the agent's choices and 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 called *intention generation* and *intention reconsideration* are defined as specific kinds of local assignments on choices. In Section 1 we introduce a static logic of time, action, and mental attitudes. In Section 2 we add the dynamic notion of local assignment to the logic of Section 1. In Section 3, we focus on two specific kinds of local assignment on choice which allow to model the processes of intention and plan generation and reconsideration.

1 A logic of time, action and mental attitudes

Let $\mathbb{N}$ be the set of non negative integers. Let $ATM^{Fact} = \{f_1, f_2, \ldots\}$ be a nonempty finite set of atoms denoting facts, and $ATM^{Act} = \{\alpha, \beta, \ldots\}$ 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. The language $\mathcal{L}$ of the logic **L** is the set of formulae defined by the following BNF:

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

where p ranges over ATM.

The three 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 $\bigcirc$ means 'φ will be true in the next state, if no event affecting the world occurs'. The operator $\bigcirc$ describes the passive/inertial evolution of the world, that is, how the world evolves over time when no event affecting it occurs. Operators of the form $[\mathtt{C}]\varphi$ have been studied in the past by [1, 2, 3]. We write $\bigcirc^n\varphi$ to indicate that the sentence φ is subject to n iterations of the modality $\bigcirc$ where $n \in \mathbb{N}$. The following abbreviation defines the concept of intention for every $\alpha \in ATM^{Act}$ and $n \in \mathbb{N}$:

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

$\mathtt{I}^n(\alpha)$ stands for 'the agent intends to do α in n steps from now'.

L-models are tuples $M = \langle H, \mathscr{B}, \mathscr{C}, \mathscr{V}\rangle$ with:

- $H = \{h, h', \ldots\}$ a nonempty set of possible histories;
- $\mathscr{B}$ and $\mathscr{C}$ 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 $\mathscr{C}(h') = \mathscr{C}(h)$;
- $\mathscr{V}$ 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, $\mathscr{C}(h)$ is the set of histories that are compatible with the agent's choices at history h. Constraints C1 and C2 express respectively that the agent's beliefs (resp. choices) are positively and negatively introspective. The satisfaction relation $\models$, between formulae and **L**-models, is defined as follows:

- $M, h(n) \models p$ iff $(h, n) \in \mathscr{V}(p)$;
- $M, h(n) \models \neg\varphi$ iff not $M, h(n) \models \varphi$;
- $M, h(n) \models \varphi \vee \psi$ iff $M, h(n) \models \varphi$ or $M, h(n) \models \psi$;
- $M, h(n) \models \bigcirc\varphi$ iff $M, h(n+1) \models \varphi$;
- $M, h(n) \models [\mathtt{C}]\varphi$ iff $M, h'(n) \models \varphi$ for all $h' \in \mathscr{C}(h)$;
- $M, h(n) \models [\mathtt{B}]\varphi$ iff $M, h'(n) \models \varphi$ for all $h' \in \mathscr{B}(h)$.

Theorem 1 *L is completely axiomatized by the following principles:*

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

2 Local assignments

We write ASG to denote the set of all partial functions σ with signature $(ATM \times \mathbb{N}) \rightarrow \mathcal{L}$. 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 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), 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 $\{(p_1,n_1) \stackrel{B}{\mapsto} \varphi_1, \ldots, (p_m,n_m) \stackrel{B}{\mapsto} \varphi_m\}$, and analogously for σ_C and σ_W. Suppose $\Sigma = (\sigma_B,\sigma_C,\sigma_W)$, we define the corresponding tuple $\Uparrow\Sigma = (\Uparrow\sigma_B, \Uparrow\sigma_C, \Uparrow\sigma_W)$ as follows: for every $n \in \mathbb{N}$ and $p \in ATM$, $\Uparrow\sigma_B(p,n) = \sigma_B(p,n{+}1)$, $\Uparrow\sigma_C(p,n) = \sigma_C(p,n{+}1)$, and $\Uparrow\sigma_W(p,n) = \sigma_W(p,n{+}1)$.

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

The language $\mathcal{L}^+$ of the logic $\mathbf{L}^+$ is defined by the following BNF:
$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\mathtt{B}]\varphi \mid [\mathtt{C}]\varphi \mid \bigcirc\varphi \mid [\Sigma{:}W]\varphi \mid [\Sigma{:}B]\varphi \mid [\Sigma{:}C]\varphi$
where p ranges over ATM and Σ ranges over $CASG$.

The formulae $[\Sigma{:}W]\varphi$, $[\Sigma{:}B]\varphi$ and $[\Sigma{:}C]\varphi$ mean respectively: φ holds in the physical world/in the context of the agent's beliefs/in the context of the agent's choices after the occurrence of the event Σ. For every $\mathbf{L}$-model M, we define the model M_n^Σ which results from the update of M at the time point n by the complex assignment Σ.

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

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). 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, 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 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.

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

The satisfaction relation between formulae in $\mathbf{L}^+$ and $\mathbf{L}$-models, is defined by the conditions of Section 1 plus:

- $M,h(n) \models [\Sigma{:}W]\varphi$ iff $M_n^\Sigma, h_W(n{+}1) \models \varphi$;
- $M,h(n) \models [\Sigma{:}B]\varphi$ iff $M_n^\Sigma, h_B(n{+}1) \models \varphi$;
- $M,h(n) \models [\Sigma{:}C]\varphi$ iff $M_n^\Sigma, h_C(n{+}1) \models \varphi$.

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

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

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

3 Intention and plan dynamics

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* an intention to do an action α n steps from now, noted $rec(\alpha,n)$.

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

The following are $\mathbf{L}^+$-theorems which highlight some interesting properties of intention generation and intention reconsideration.

(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 $m \neq n$

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 α $n{+}1$ steps from now, in the physical world the agent does not intend to do α n steps from now. In the definition of the truth condition of the operators $[\Sigma{:}C]$ we have supposed that the occurrence of a local assignment takes time (one time unit). Consequently, also the processes of generating/reconsidering an intention take 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. 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 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.

REFERENCES

[1] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[2] E. Lorini and A. Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.

[3] A. S. Rao and M. P. Georgeff. Modelling rational agents within a BDI-architecture. In *Proc. of KR'91*, pages 473–484, 1991. Morgan Kaufmann.