

# $CTL_{AgentSpeak(L)}$ : a Specification Language for Agent Programs

Alejandro Guerra-Hernández<sup>1</sup>, José Martín Castro-Manzano<sup>1</sup>, and Amal El-Fallah-Seghrouchni<sup>2</sup>

<sup>1</sup> Departamento de Inteligencia Artificial  
Universidad Veracruzana  
Facultad de Física e Inteligencia Artificial  
Sebastián Camacho No. 5, Xalapa, Ver., México, 91000  
`aguerra@uv.mx`, `e.f.s.s@hotmail.com`  
<sup>2</sup> Laboratoire d'Informatique de Paris 6  
Université Pierre et Marie Curie  
Avenue du Président Kennedy, Paris, France, 75016  
`Amal.Elfallah@lip6.fr`

**Abstract.** This work introduces  $CTL_{AgentSpeak(L)}$ , a logic to specify and verify expected properties of rational agents implemented in the  $AgentSpeak(L)$  agent oriented programming language. Our approach is similar to the classic  $BDICTL$  modal logic, used to reason about agents modelled in terms of belief (BEL), desires (DES), intentions (INTEND). A new interpretation for the temporal operators in  $CTL$ : next ( $\bigcirc$ ), eventually ( $\diamond$ ), until(U), inevitable(A), etc., is proposed in terms of the transition system induced by the operational semantics of  $AgentSpeak(L)$ . The main contribution of the approach is a better understanding of the relation between the programming language and its logical specification, enabling us to proof expected or desired properties for any agent programmed in  $AgentSpeak(L)$ , e.g., commitment strategies.

## 1 Introduction

The theory of practical reasoning proposed by Bratman [3], expounds the philosophical foundation for the computational approaches to rational agency, known as BDI (Belief-Desire-Intention) systems. This theory is innovative because it does not reduce intentions to some combination of beliefs and desires, but indeed it assumes that intentions are composed by hierarchical, partial plans. Such assumption explains better temporal aspects of practical reasoning as future intentions, persistence, reconsideration and commitment. Different multi-modal BDI logics [14, 16, 17] have been proposed to formally characterize the rational behavior of such agents, in terms of the properties of the intentional attitudes and the relations among them, e.g., it is rational to intend desires that are believed possible; and temporal changes of these mental attitudes, e.g., commitment strategies define when it is rational to drop intentions. Due to their expressiveness, these logics are used to reason about the agents properties; but, because of their computational cost, they are not used to program them.

Agent oriented programming languages, such as *AgentSpeak(L)* [13], have been proposed to reduce the gap between theory (logical specification) and practice (implementation) of rational agents. Even when this programming language has a well defined operational semantics, the verification of rational properties of the programmed agents is not evident, since intentional and time modalities are abandoned for the sake of efficiency. In order to reason about such properties, we propose  $CTL_{AgentSpeak(L)}$  as a logic for specification and verification of *AgentSpeak(L)* agents. The approach is similar to the classic  $BDICTL$  [14] logic, defined as a  $B^{KD45}D^{KD}I^{KD}$  modal system, with temporal operators: next ( $\bigcirc$ ), eventually ( $\diamond$ ), until( $U$ ), inevitable( $A$ ), etc., defined after the computational tree logic (CTL) [6].

Our main contribution is the definition of the semantics of the  $CTL$  temporal operators in terms of a Kripke structure, produced by a transition system defining the operational semantics of *AgentSpeak(L)*. The semantics of the intentional operators is adopted from the work of Bordini et al. [1]. As a result, the semantics of  $CTL_{AgentSpeak(L)}$  is grounded in the operational semantics of programming language. In this way, we can prove if any agent programmed in *AgentSpeak(L)* satisfies certain properties expressed in the logical specification. It is important to notice that our problem is different from that of model checking in the following sense: in model checking the problem consists in verifying if certain property holds in certain state in certain agent, while our work deals with verifying that certain general properties hold for any agent. The approach is exemplified verifying the commitment strategies for *AgentSpeak(L)*.

The paper is organized as follows: Section 2 exemplifies the specification of rational properties in  $BDICTL$  [14] with the definition of commitment strategies. Section 3 introduces the syntax and the semantics of the *AgentSpeak(L)* agent oriented programming language. Section 4 presents the main contribution of the paper: a logic framework to reason about *AgentSpeak(L)* agents, with semantics based on the operational semantics of the programming language. Section 5 shows how the commitment strategies introduced in section 2 can be verified for *AgentSpeak(L)*. Section 6 offers concluding remarks and discusses future work.

## 2 Commitment

As mentioned, different computational theories have been proposed to capture the theory of Bratman [3] on intentions, plans and practical reasoning. The foundational work of Cohen and Levesque [4], for example, defined intention as a combination of belief and desire based on the concept of persistent goal. A critical analysis of this theory [15] showed that the theory of Cohen and Levesque failed to capture important aspects of commitment. Alternatively, commitment has been approached as a process of maintenance and revision of intentions, relating current and future intentions.

Different types of commitment strategies define different types of agents. Three of them have been extensively studied in the context of  $BDICTL$  [14], where  $CTL$  denotes computational tree logic [6], the well known temporal logic:

- **Blind commitment.** An agent intending that inevitably (A) eventually ( $\diamond$ ) is the case that  $\phi$ , inevitably maintains his intentions until (U) he actually believes  $\phi$ :

$$\text{INTEND}(A\diamond\phi) \implies A(\text{INTEND}(A\diamond\phi) \text{ U } \text{BEL}(\phi)) \quad (1)$$

- **Single-minded commitment.** An agent maintains his intentions as long as he believes they are achieved or optionally (E) eventually achievable:

$$\text{INTEND}(A\diamond\phi) \implies A(\text{INTEND}(A\diamond\phi) \text{ U } (\text{BEL}(\phi) \vee \neg\text{BEL}(E\diamond\phi))) \quad (2)$$

- **Open-minded commitment.** An agent maintains his intentions as long as they are achieved or they are still desired:

$$\text{INTEND}(A\diamond\phi) \implies A(\text{INTEND}(A\diamond\phi) \text{ U } (\text{BEL}(\phi) \vee \neg\text{DES}(A\diamond\phi))) \quad (3)$$

For example, a blind agent intending eventually to go to Paris, will maintain his intention, for any possible course of action (inevitable), until he believes he is going to Paris. A single-minded agent can drop such intention if he believes it is not possible any more to go to Paris. An open-minded agent can drop the intention if he does not desire anymore going to Paris. An interesting question is what kind of commitment strategy is followed by *AgentSpeak(L)* agents?

Furthermore, how can we relate commitment with policy-based reconsideration in *AgentSpeak(L)*? Bratman argues that there are three cases of reconsideration ([3], pp. 60–75) in practical reasoning. Non-reflective reconsideration has short effects, while deliberative one is very expensive. Policy-based reconsideration is a compromise between impact and cost. Obviously, if an agent is blindly committed, we can not talk about any form of reconsideration. But if the an agent is single-minded, then we could approach policy-based reconsideration through intentional learning [7–10]. In this way we would reconcile a relevant aspect of the computational theories of BDI agency (commitment) with its philosophical foundation (reconsideration *à la* Bratman).

### 3 *AgentSpeak(L)*

In this section, the syntax and semantics of *AgentSpeak(L)* [13], as defined for its interpreter Jason [2], are introduced. The operational semantics of the language is given in terms of a transition system that, being a Kripke structure, is used to define the semantic of the temporal operators in  $CTL_{AgentSpeak(L)}$  in the next section; the BDI operators will also be defined in terms of the structures supporting the operational semantics of *AgentSpeak(L)*

### 3.1 Syntax

The syntax of *AgentSpeak(L)* is shown in table 1. As usual, an agent *ag* is formed by a set of plans *ps* and beliefs *bs* (grounded literals). Each plan has the form *triggerEvent : context*  $\leftarrow$  *body*. The context of a plan is a literal or a conjunction of them. A non empty plan body is a finite sequence of actions, goals (achieve or test an atomic formula), or beliefs updates (addition or deletion).  $\top$  denotes empty elements, e.g., plan bodies, contexts, intentions. The trigger events are updates (addition or deletion) of beliefs or goals.

$ag ::= bs \ ps$	$at ::= P(t_1, \dots, t_n) \ (n \geq 0)$
$bs ::= b_1 \dots b_n$	$a ::= A(t_1, \dots, t_n) \ (n \geq 0)$
$ps ::= p_1 \dots p_n$	$g ::= !at \   \ ?at$
$p ::= te : ct \leftarrow h$	$u ::= +b \   \ -b$
$te ::= +at \   \ -at \   \ +g \   \ -g$	
$ct ::= ct_1 \   \ \top$	
$ct_1 ::= at \   \ \neg at \   \ ct_1 \wedge ct_1$	
$h ::= h_1; \top \   \ \top$	
$h_1 ::= a \   \ g \   \ u \   \ h_1; h_1$	

**Table 1.** Syntax of *AgentSpeak(L)*. Adapted from Bordini et al. [2]

### 3.2 Operational semantics

The operational semantics of *AgentSpeak(L)* is defined as a transition system between configurations  $\langle ag, C, M, T, s \rangle$ , where:

- *ag* is an agent program formed by beliefs *bs* and plans *ps*.
- An agent circumstance *C* is a tuple  $\langle I, E, A \rangle$  where *I* is the set of intentions  $\{i, i', \dots, n\}$  s.t. *i*  $\in$  *I* is a stack of partially instantiated plans  $p \in ps$ ; *E* is a set of events  $\{\langle te, i \rangle, \langle te', i' \rangle, \dots, n\}$ , s.t. *te* is a *triggerEvent* and each *i* is an intention (internal event) or an empty intention  $\top$  (external event); and *A* is a set of actions to be performed by the agent in the environment.
- *M* is a tuple  $\langle In, Out, SI \rangle$  that works as a *mailbox*, where *In* is the mailbox of the agent, *Out* is a list of messages to be delivered by the agent and *SI* is a register of suspended intentions (intentions that wait for an answer message). It is not relevant for the purposes of this paper.
- *T* is a tuple  $\langle R, Ap, \iota, \epsilon, \rho \rangle$  that registers temporal information: *R* is the set of relevant plans given certain *triggerEvent*; *Ap* is the set of applicable plans (the subset of *R* s.t.  $bs \models ctx$ );  $\iota, \epsilon$  y  $\rho$  register, respectively, the intention, the event and the current plan during an agent execution.
- The configuration label  $s \in \{SelEv, RelPl, AppPl, SelAppl, SelInt, AddIM, ExecInt, ClrInt, ProcMsg\}$  indicates the current step in the reasoning cycle of the agent.

Transitions are defined in terms of semantic rules of the form:

$$(\text{rule id}) \quad \frac{\text{cond}}{C \rightarrow C'}$$

where  $C = \langle ag, C, M, T, s \rangle$  is an *AgentSpeak(L)* configuration that can be transformed to a new one  $C'$ , if the conditions *cond* hold. Table 2 shows the semantic rules that are relevant for the purposes of this paper (communication processing rules are omitted for simplicity).

(SelEv <sub>1</sub> )	$\frac{S_E(C_E) = \langle te, i \rangle}{\langle ag, C, M, T, SelEv \rangle \longrightarrow \langle ag, C', M, T', RelPl \rangle}$	s.t. $C'_E = C_E \setminus \{\langle te, i \rangle\}$ $T'_\epsilon = \langle te, i \rangle$
(Rel <sub>1</sub> )	$\frac{T_\epsilon = \langle te, i \rangle, RelPlans(ag_{ps}, te) \neq \{\}}{\langle ag, C, M, T, RelPl \rangle \longrightarrow \langle ag, C, M, T', AppPl \rangle}$	s.t. $T'_R = RelPlans(ag_{ps}, te)$
(Rel <sub>2</sub> )	$\frac{RelPlans(ps, te) = \{\}}{\langle ag, C, M, T, RelPl \rangle \longrightarrow \langle ag, C, M, T, SelEv \rangle}$	
(App <sub>1</sub> )	$\frac{AppPlans(ag_{bs}, T_R) \neq \{\}}{\langle ag, C, M, T, AppPl \rangle \longrightarrow \langle ag, C, M, T', SelApp \rangle}$	s.t. $T'_{Ap} = AppPlans(ag_{bs}, T_R)$
(SelApp <sub>1</sub> )	$\frac{S_O(T_{Ap}) = (p, \theta)}{\langle ag, C, M, T, SelApp \rangle \longrightarrow \langle ag, C, M, T', AddIM \rangle}$	s.t. $T'_\rho = (p, \theta)$
(ExtEv)	$\frac{T_\epsilon = \langle te, \top \rangle, T_\rho = (p, \theta)}{\langle ag, C, M, T, AddIM \rangle \longrightarrow \langle ag, C', M, T, SelInt \rangle}$	s.t. $C'_I = C_I \cup \{p\theta\}$
(SelInt <sub>1</sub> )	$\frac{C_I \neq \{\}, S_I(C_I) = i}{\langle ag, C, M, T, SelInt \rangle \longrightarrow \langle ag, C, M, T', ExecInt \rangle}$	s.t. $T'_i = i$
(SelInt <sub>2</sub> )	$\frac{C_I = \{\}}{\langle ag, C, M, T, SelInt \rangle \longrightarrow \langle ag, C, M, T, ProcMsg \rangle}$	
(AchvG <sub>1</sub> )	$\frac{T_i = i[\text{head} \leftarrow !at; h]}{\langle ag, C, M, T, ExecInt \rangle \longrightarrow \langle ag, C', M, T, ProcMsg \rangle}$	s.t. $C'_E = C_E \cup \{\langle +!at, T_i \rangle\}$ $C'_I = C_I \setminus \{T_i\}$
(ClrInt <sub>1</sub> )	$\frac{T_i = [\text{head} \leftarrow \top]}{\langle ag, C, M, T, ClrInt \rangle \longrightarrow \langle ag, C', M, T, ProcMsg \rangle}$	s.t. $C'_I = C_I \setminus \{T_i\}$
(ClrInt <sub>2</sub> )	$\frac{T_i = i[\text{head} \leftarrow \top]}{\langle ag, C, M, T, ClrInt \rangle \longrightarrow \langle ag, C', M, T, ClrInt \rangle}$	s.t. $C'_I = (C_I \setminus \{T_i\}) \cup$ $\{k[(\text{head}' \leftarrow h)\theta]\}$ if $i = k[\text{head}' \leftarrow g; h]$ and $g\theta = TrEv(\text{head})$
(ClrInt <sub>3</sub> )	$\frac{T_i \neq [\text{head} \leftarrow \top] \wedge T_i \neq i[\text{head} \leftarrow \top]}{\langle ag, C, M, T, ClrInt \rangle \longrightarrow \langle ag, C, M, T, ProcMsg \rangle}$	

**Table 2.** Operational semantics rules of *AgentSpeak(L)*.

The reasoning cycle of an *AgentSpeak(L)* agent starts processing messages and updating perception ( $s = ProcMsg$ ). This adds events to  $C_E$  ( $C_\alpha$  denotes the element  $\alpha$  of circumstance  $C$ , the same for other element of configurations)

to be processed by the agent. One of these events is selected ( $\mathcal{S}_E(C_E) = \langle te, i \rangle$ ), and relevant and applicable plans are generated using the following definitions:

**Definition 1 (Relevant plans)** *Given a set of plans  $ag_{ps}$ , the subset of relevant plans for a selected event  $\langle te, i \rangle \in C_E$ , is defined as:*

$$RelPlans(ps, te) = \{(p, \theta) | p \in ps \wedge \theta = mgu(te, TrEv(p))\}$$

where  $TrEv(te' : ctxt \leftarrow h) = te'$  gets the trigger event  $te$  of a plan,  $C_E$  denotes the events  $E$  in a given circumstance  $C$ , and  $mgu$  is the most general unifier.

**Definition 2 (Applicable plans)** *Given a set of relevant plans  $T_R$  and a set of beliefs  $ag_{bs}$ , the set of applicable plans is defined as:*

$$AppPlans(bs, R) = \{(p, \theta\theta') | (p, \theta) \in R \wedge \theta' \text{ is s.t. } bs \models Ctxt(p)\theta\theta'\}$$

where  $\theta'$  is the substitution computed when verifying if the context of relevant plan  $p$  ( $Ctxt(p)$ ), affected by its relevant substitution  $\theta$ , is a logical consequence of the beliefs of the agent  $bs$ .

Then the agent proceeds selecting a relevant plan to form an intention ( $Appl_1$  and  $SelAppl$ ) or, if no relevant plans were found, selecting an intention to be executed ( $Appl_2$  and  $SelInt_1$ ). The execution of an intention changes the environment and the mental attitudes of the agent (including the abandoning of accomplished intentions in  $ClrInt$ ).  $ProcMsg$  generates the new events, and so on. Figure 1 shows the transition system induced by these semantic rules. States are labeled with possible values for  $s$ . Transitions correspond to the semantic rules identifiers. The initial state is  $s = ProcMsg$ .

Although the operational semantics of  $AgentSpeak(L)$  clearly defines the practical reasoning performed by an agent, it is difficult to prove BDI properties, such as the commitment strategies, for any given agent. This is due to the abandon of intentional and temporal modalities in  $AgentSpeak(L)$ , the main reason to propose  $CTL_{AgentSpeak(L)}$ .

#### 4 $CTL_{AgentSpeak(L)}$

$CTL_{AgentSpeak(L)}$  may be seen as an instance of  $BDI_{CTL}$ . Similar approaches have been explored for other instances of agent oriented programming languages, e.g, a simplified version of 3APL [5]. The idea is to define the semantics of temporal and intentional operators in terms of the operational semantics of the programming language. This grounds the semantics to reason about particular kinds of agents, in this case  $AgentSpeak(L)$  agents. Once this is done, we can use the logic to reason about general properties of such agents.

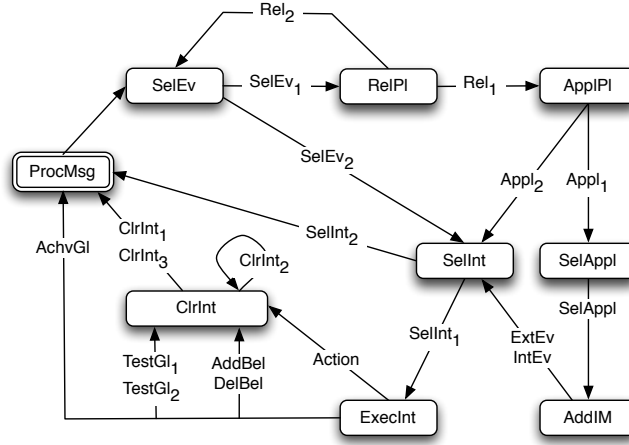


Fig. 1. The interpreter for  $AgentSpeak(L)$  as a transition system.

#### 4.1 Syntax

Formulae for intentional and temporal operators are required:

**Definition 3 (BDI syntax)** *If  $\phi$  is an atomic formula in  $AgentSpeak(L)$ , then  $\phi$ ,  $BEL(\phi)$ ,  $DES(\phi)$ ,  $INTEND(\phi)$  are  $CTL_{AgentSpeak(L)}$  well formed formulae (BDI formulae).*

**Definition 4 (Temporal syntax)** *Temporal formulae are divided, as usual, in state and path formulae. State formulae are defined as:*

- s1 *Each well formed BDI formula is a state formula.*
- s2 *If  $\phi$  and  $\psi$  are state formulae,  $\phi \wedge \psi$  and  $\neg\phi$  are state formulae.*
- s3 *If  $\phi$  is a path formula, then  $E\phi$  and  $A\phi$  are state formulae.*

*Path formulae are defined as:*

- p1 *Each state formula is also a path formula.*
- p2 *If  $\phi$  and  $\psi$  are path formulae, then  $\neg\phi$ ,  $\phi \wedge \psi$ ,  $\bigcirc\phi$ ,  $\diamond\phi$  and  $\phi \text{ U } \psi$  are path formulae.*

For example,  $INTEND(A\diamond go(paris)) \text{ U } \neg BEL(go(paris, summer))$ , expressing that the agent will intend inevitably (A) for every course of action, eventually ( $\diamond$ ) going to Paris until (U) he does not believe go to Paris in summer, is a well formed formula.

## 4.2 Semantics

The semantics of the intentional operators BEL, DES and INTEND is adopted from Bordini et al. [1]. First an auxiliary function for getting the atoms ( $at$ ) subject of an achieve goal ( $+!$ ) in a given intention, is defined: required:

$$\begin{aligned} agoals(\top) &= \{\}, \\ agoals(i[p]) &= \begin{cases} \{at\} \cup agoals(i) & \text{if } p = +!at : ct \leftarrow h, \\ agoals(i) & \text{otherwise} \end{cases} \end{aligned}$$

**Definition 5 (BDI semantics)** *The semantics of the BEL, DES, and INTEND operators for a given agent  $ag$  and its circumstance  $C$  is:*

$$\begin{aligned} BEL_{\langle ag, C \rangle}(\phi) &\equiv ag_{bs} \models \phi \\ INTEND_{\langle ag, C \rangle}(\phi) &\equiv \phi \in \bigcup_{i \in C_I} agoals(i) \vee \bigcup_{\langle te, i \rangle \in C_E} agoals(i) \\ DES_{\langle ag, C \rangle}(\phi) &\equiv \langle +!\phi, i \rangle \in C_E \vee INTEND(\phi) \end{aligned}$$

If the agent  $ag$  and his circumstance  $C$  are explicit, we simply write  $BEL(\phi)$ ,  $DES(\phi)$ , and  $INTEND(\phi)$ . So an agent  $ag$  is said to believe the atomic formula  $\phi$ , if  $\phi$  is a logical consequence of the beliefs  $bs$  of  $ag$ . An agent is said to intend the atomic formula  $\phi$ , if  $\phi$  is the subject of an achieve goal in the active intentions of the agent ( $C_I$ ) or in his suspended intentions associated to events to be processed ( $C_E$ ). An agent is said to desire the atomic formula  $\phi$ , if there is an event in  $C_E$  which trigger is an achieve goal about  $\phi$  or if  $\phi$  is intended.

The semantics of the temporal operators: next( $\bigcirc$ ), eventually ( $\diamond$ ), and until ( $\mathbf{U}$ ), as well as the path quantifier inevitable ( $\mathbf{A}$ ), required a Kripke structure  $\langle S, R, L \rangle$  where  $S$  is a set of states,  $R$  is a serial relation on  $S \times S$  and  $L$  is a labelling or a valuation function for formulae in the states. The transition system of *AgentSpeak(L)* induce a Kripke structure:

**Definition 6 (*AgentSpeak(L)* Kripke structure)**  $K = \langle S, R, V \rangle$  is a Kripke structure specified by the transition system ( $\Gamma$ ) defining the *AgentSpeak(L)* operational semantics rules:

- $S$  is a set of agent configurations  $\langle ag, C, M, T, s \rangle$ .
- $R \subseteq S^2$  is a serial relation s.t. for all  $(c_i, c_j) \in R$ ,  $(c_i, c_j) \in \Gamma$  or  $c_i = c_j$ .
- $V$  is a valuation function over the intentional and temporal operators, defined after their semantics (see definitions 5 and 7), e.g.,  $V_{BEL}(c, \phi) \equiv BEL(phi)$  at the configuration  $c = \langle ag, C, M, T, s \rangle$ , etc.

As usual,  $x = c_0, \dots, c_n$  denotes a path in the Kripke structure, i.e., a sequence of configurations s.t. for all  $c_i \in S$ ,  $(c_i, c_{i+1}) \in R$ . The expression  $x^i$  denotes the suffix of path  $x$  starting at configuration  $c_i$ .



**Definition 7 (Temporal semantics)** *The semantic of the state formulae is defined for a given current configuration  $c_i \in K$ :*

$$\begin{aligned}
K, c_i \models \text{BEL}(\phi) &\Leftrightarrow \phi \in V_{\text{BEL}}(c_i, \phi) \\
K, c_i \models \text{INTEND}(\phi) &\Leftrightarrow \phi \in V_{\text{INTEND}}(c_i, \phi) \\
K, c_i \models \text{DES}(\phi) &\Leftrightarrow \phi \in V_{\text{DES}}(c_i, \phi) \\
K, c_i \models \text{E}\phi &\Leftrightarrow \exists x^i \exists c_{j \geq i} \in x^i \text{ s.t. } K, c_j \models \phi \\
K, c_i \models \text{A}\phi &\Leftrightarrow \forall x^i \exists c_{j \geq i} \in x^i \text{ s.t. } K, c_j \models \phi
\end{aligned}$$

*The semantic of the path formulae is defined as follows:*

$$\begin{aligned}
K, c_i \models \phi &\Leftrightarrow K, x^i \models \phi, \text{ where } \phi \text{ is a state formula} \\
K, c_i \models \bigcirc \phi &\Leftrightarrow K, x^{i+1} \models \phi \\
K, c_i \models \diamond \phi &\Leftrightarrow \exists c_{j \geq i} \in x^i \text{ s.t. } K, c_j \models \phi \\
K, c_i \models \phi \text{ U } \psi &\Leftrightarrow (\exists c_{k \geq i} \text{ s.t. } K, x^k \models \psi \wedge \forall c_{i \leq j < k} \quad K, x^j \models \phi) \\
&\vee (\exists c_{j \geq i} \quad K, x^j \models \phi).
\end{aligned}$$

Observe that the semantics of until corresponds to weak until ( $\psi$  can never occur). Once satisfaction over state and path formulae has been defined, we can define satisfaction and validity over *AgentSpeak(L)* runs.

**Definition 8 (Run)** *Given an initial *AgentSpeak(L)* configuration  $c_0$ , the run  $K_\Gamma^0$  denotes the sequence of configurations  $c_0, c_1, \dots$  such that  $\forall i \geq 1, c_i = \Gamma(c_{i-1})$ .*

**Definition 9 (Satisfaction over runs)** *Given an *AgentSpeak(L)* run  $K_\Gamma^0$  the property  $\phi \in \text{CTL}_{\text{AgentSpeak(L)}}$  is satisfied if  $\forall i \geq 0, K_\Gamma^0, c_i \models \phi$ .*

**Definition 10 (Validity)** *A property  $\phi \in \text{CTL}_{\text{AgentSpeak(L)}}$  is valid if for any initial configuration  $K_\Gamma^0, c_0 \models \phi$ , e.g., if it is satisfied for any run of any agent.*

## 5 Results about commitment

Now we proceed to show some properties that expressed in the logical specification hold for any *AgentSpeak(L)* agent. We choose to verify blind and single-minded commitment strategies. In principle, open-minded commitment can be verified in a similar way to the single-minded one, but since it is related to emotional aspects of agency more than to policy-based reconsideration (see section 2), it has not been included here. First the no-infinite deferral axiom is verified. It expresses that there is no agent that maintains forever his intentions.

**Proposition 1** *An *AgentSpeak(L)* agent satisfies the axiom of no-infinite deferral:  $\text{INTEND}(\phi) \Rightarrow \text{A}\diamond(\neg\text{INTEND}(\phi))$ .*

*Proof.* Assume  $K, c_0 \models \text{INTEND}(\phi)$ , then given the definition for INTEND (definition 5), there is a plan  $p \in C_I \cup C_E$  with head  $+\!|\phi$  at  $c_0$ . The non-infinite deferral axiom expresses that for all runs  $K_I^0$  eventually this plan will be removed from  $C_I$  (active intentions) and  $C_E$  (suspended intentions). While  $p$  is being executed successfully, three runs are possible given the transition rules  $\text{ClrInt}_X \in \Gamma$  (table 2): i)  $\text{ClrInt}_3$  applies when the body of  $p$  is not empty, nothing is cleaned; ii)  $\text{ClrInt}_2$  applies when the plan  $p$ , with an empty body, is at the top of an intention  $i$ , then  $p$  is dropped from  $i$ ;  $\text{ClearInt}_1$  applies when the intention  $i$  has only a plan  $p$  with an empty body, the whole intention  $i$  is dropped. Given the finite nature of the plans (section 3.1), if everything goes right, conditions (ii) or (iii) are eventually reached. If something goes wrong with  $p$ , a failure mechanism is activated by an event of the form  $\langle -\!|\phi, i[p] \rangle$  resulting in  $p$  being dropped. Although [2] only formalizes failures in finding relevant plans ( $\text{Rel}_2$ ), which discards suspended intentions in  $C_E$ , other forms of failure detection have been considered in the context of intentional learning [7, 8]. By successful or failed execution of the plans every adopted intention is eventually dropped.  $\square$

**Proposition 2** *An AgentSpeak(L) agent does not satisfy the axiom of blind commitment.*

*Proof.* Given that *AgentSpeak(L)* agents satisfy the no-infinite deferral property (proposition 1) and the weak until semantics, the blind commitment axiom is reduced to (a similar reduction is used in Rao et al. [14]):

$$\text{INTEND}(\text{A}\diamond\phi) \Rightarrow \text{A}\diamond\text{BEL}(\phi)$$

Consider an initial configuration  $c_0$  s.t.  $ag = \langle bs, ps \rangle$  where  $bs = \{\}$  and  $ps = \{+\!|b(t_1) : \top \leftarrow p(t_2). \quad +\!|p(t_2) : \top \leftarrow +\!|b(t_3).\}$ . Suppose that from perception of the environment  $a$  is added  $ag_{bs} = \{b(t_1)\}$ . An event is generated by this belief update, so that  $C_E = \{\langle +\!|b(t_1), \top \rangle\}$ . Then following the semantic rules defining  $\Gamma$ ,  $\text{SelEv}_1$ ,  $\text{Rel}_1$ ,  $\text{AppPl}_1$ , are applied obtaining a configuration where  $C_I = \{\langle +\!|b(t_1) : \top \leftarrow !p(t_2).\rangle\}$  and  $C_E = \{\}$ . Then proceeding with the rules  $\text{SelAppl}$ ,  $\text{ExtEv}$ ,  $\text{SelInt}_1$ ,  $\text{AchvGl}$  a configuration where  $C_E \langle +\!|p(t_2), +\!|b(t_1) : \top \leftarrow \top \rangle$ ,  $C_I = \{\}$  is obtained. At this configuration  $c'$ ,  $K, c' \models \text{DES}(p(t_2))$  (Definition. 5). If we apply then  $\text{SelEv}_1$ ,  $\text{Rel}_1$ ,  $\text{AppPl}_1$ ,  $\text{SelAppl}$ , a configuration where  $C_I = \{\langle +\!|p(t_2) : \top \leftarrow +\!|b(t_3).\rangle\}$  and  $C_E = \{\}$ , is obtained. In this configuration  $c''$   $K, c'' \models \text{INTEND}(p(t_2))$  (Definition. 5). Then proceeding with  $\text{IntEv}$ ,  $\text{SelInt}_1$ ,  $\text{AddBel}$  gets  $C_E = \langle +\!|b(t_3), \top \rangle$ ,  $ag_{bs} = \{b(t_1)\}$  and  $C_I = \{\langle +\!|b(t_1) : \top \leftarrow \top \quad \ddagger +\!|p(t_2) : \top \leftarrow \top \rangle\}$  and  $bs = \{b(t_1), b(t_3)\}$ . The intention about  $p(t_2)$  is maintained. Observe that the plan bodies in the intention are empty, so the  $\text{ClrInt}$  rules will discard the whole intention, so that at the next configuration  $c'''$ ,  $K, c''' \models \neg\text{INTEND}(p(t_2))$  and  $K, c''' \models \neg\text{BEL}(p(t_2))$ . By counter-example  $\text{INTEND}(\text{A}\diamond\phi) \Rightarrow (\text{A}\diamond(\text{BEL}(\phi)))$  is not valid.  $\square$

In fact, our agents do neither satisfy the extended blind commitment axiom (eq. 1), since the agent did not keep its intention about  $p(t_2)$  until she believed it.

This reasoning is similar to the demonstration of intention-belief incompleteness (AT2) for *AgentSpeak(L)* [1].

**Proposition 3** *AgentSpeak(L) agents satisfy a limited single-minded commitment:  $\text{INTEND}(A\Diamond\phi) \implies A(\text{INTEND}(A\Diamond\phi) \cup \neg\text{BEL}(E\Diamond\phi))$ .*

*Proof.* This case is similar to the no-infinite deferral demonstration. Assume the agent  $\text{INTEND}(A\Diamond\phi)$  at  $c_0$ , then there is a plan  $p \in C_I \cup C_E$  with head  $+!\phi$  at  $c_0$ . If there is a configuration  $c_k \geq 0$  where  $\neg\text{BEL}(E\Diamond\phi)$  (the weak until definition has been adopted), then  $K, x^0, \dots, x^k \models \text{INTEND}(A\Diamond\phi)$ . Following the no-infinite-deferral demonstration, in the failure cases the agent will eventually satisfy  $\bigcirc\neg\text{INTEND}(\phi)$  because  $\text{Rel}_2$  which means that for an event  $\langle te, i[+!\phi : c \leftarrow h.] \rangle$  there were not relevant plans and the associated intention will be discarded, e.g., there is not a path of configurations where eventually  $\phi$ , so that it is rational to drop  $\text{INTEND}_{(ag, C)}(\phi)$ . The case of no-infinite deferral by successful execution of intentions covers the second condition of the weak until  $\neg\text{BEL}(E\Diamond\phi)$  does not occur.  $\square$

This is a limited form of single-minded commitment because  $\neg\text{BEL}(E\Diamond\phi)$  is not represented explicitly by the agent. In fact, the agent can not continue intending  $\phi$  because there are no plans to do it and the full intention fails. It is also limited because of intention-belief incompleteness that can be avoided dropping the close world assumption [1]; or using the intentional and temporal definitions for studying the necessary conditions in the operational semantics and definition of the agents to warranty the expected properties of intentions, e.g., in the case of intentions, what does it mean in terms of *AgentSpeak(L)* to be equivalent to a *KD* modal system?

Given that, *AgentSpeak(L)* agents are not blind committed, intentional learning [7–10] provides a third alternative approach to achieve a full single-minded commitment. The idea is that the agents can learn, in the same way they learn the successful adoption of plans as intentions, the reasons behind a plan adoption that lead to an intention failure. In this way a policy-based reconsideration can be approached.

## 6 Conclusion

We have extended the methodology proposed by Bordini et al. [1] to reason about *AgentSpeak(L)* agents. Then we proved that any *AgentSpeak(L)* agent is not blindly committed, but follows a limited form of single-minded commitment. The main limitations for these agents are intention-belief incompleteness and the lack of an explicit representation for abandoning reasons. Guerra et al. [8, 9] have suggested that intentional learning provides a solution for the latter, enabling a policy-based reconsideration.

Interestingly, the degree of boldness and cautiousness for a given agent is something hard to define. It is well known [11] that in dynamic environments a very cautious agent performs better than a bold one; and inversely, in static environments boldness pays better. The relevance of learning intentionally is

that the right degree of cautionness is learned by the agents, instead of being established once and forever by the programmers. An extended *AgentSpeak(L)* operational semantics that deals with intentional learning, for both incremental and batch inductive methods, has been proposed [10]. So, it is possible to arrive to a full theory of commitment and intentional learning, using the techniques presented here. We are currently experimenting these ideas.

As the example of commitment, reconsideration and learning illustrates, the properties verified in this paper are not arbitrary ones. Proving these properties using  $CTL_{AgentSpeak(L)}$ , prove that they hold for any *AgentSpeak(L)* agent. This also illustrates the relevance of the specification language proposed in this paper, to bring *AgentSpeak(L)* closer to its philosophical foundation and to extend our computational theories of practical reasoning.

**Acknowledgments.** The first and third authors are supported by Conacyt CB-2007-1 (project 78910) funding for this research. The second author is supported by Conacyt scholarship 214783.

## References

1. Bordini, R.H., Moreira, Á.F.: Proving BDI properties of agent-oriented programming languages. *Annals of Mathematics and Artificial Intelligence* 42, 197–226 (2004)
2. Bordini, R.H., Hübner, J.F., Wooldridge, M.: *Programming Multi-Agent Systems in AgentSpeak using Jason*. Wiley, England (2007)
3. Bratman, M.: *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge (1987)
4. Cohen, P., Levesque, H.: Intention is choice with commitment. *Artificial Intelligence* 42(3), 213–261 (1990)
5. Dastani, M., van Riemsdijk, M.B., Meyer, J.C.: A grounded specification language for agent programs. In: *AAMAS '07*. ACM, New York, NY, pp. 1–8 (2007)
6. Emerson, A.: Temporal and modal logic. In: van Leeuwen, J., (ed.) *Handbook of Theoretical Computer Science*, pp. 996–1072. Elsevier, Amsterdam (1990)
7. Guerra-Hernández, A., El-Fallah-Seghrouchni, A., Soldano, H.: Learning in BDI Multi-agent Systems. In: Dix, J., Leite, J. (eds.) *CLIMA IV*. LNCS, vol. 3259, pp. 218–233. Springer, Heidelberg (2004)
8. Guerra-Hernández, A., Ortíz-Hernández, G.: Toward BDI sapient agents: Learning intentionally. In: Mayorga, R.V., Perlovsky, L.I. (eds.) *Toward Artificial Sapience: Principles and Methods for Wise Systems*, pp. 77–91. Springer, London (2008)
9. Guerra-Hernández, A., Castro-Manzano, J.M., El-Fallah-Seghrouchni, A.: Toward an AgentSpeak(L) theory of commitment and intentional learning. In: Gelbuc, A., Morales, E.F. (eds.), *MICAI 2008*. LNCS, vol. 5317, pp. 848–858, Springer-Verlag, Berlin Heidelberg (2008)
10. Guerra-Hernández, A., Ortíz-Hernández, G., Luna-Ramírez, W.A.: Jason smiles: Incremental BDI MAS learning. In: *MICAI 2007 Special Session*, IEEE, Los Alamitos (In press)
11. Kinny, D., Georgeff, M.P.: Commitment and effectiveness of situated agents. In: *Proceedings of the twelfth international joint conference on artificial intelligence (IJCAI-91)*, Sydney, Australia (1991)

12. Rao, A.S., Georgeff, M.P.: Modelling Rational Agents within a BDI-Architecture. In: Huhns, M.N., Singh, M.P., (eds.) *Readings in Agents*, pp. 317–328. Morgan Kaufmann (1998)
13. Rao, A.S.: AgentSpeak(L): BDI agents speak out in a logical computable language. In: de Velde, W.V., Perram, J.W. (eds.) *MAAMAW. LNCS*, vol. 1038, pp. 42–55. Springer, Heidelberg (1996)
14. Rao, A.S., Georgeff, M.P.: Decision procedures for BDI logics. *Journal of Logic and Computation* 8(3), pp. 293–342 (1998)
15. Singh, M.P.: A critical examination of the Cohen-Levesque Theory of Intentions. In: *Proceedings of the European Conference on Artificial Intelligence* (1992).
16. Singh, M.P., Rao, A.S., Georgeff, M.P.: Formal Methods in DAI: Logic-Based Representation and Reasoning. In: *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pp. 331–376. MIT Press, Cambridge (1999)
17. Wooldridge, M.: *Reasoning about Rational Agents*. MIT Press, Cambridge (2000)