Consistency and Soundness for a Defeasible Logic of Intention

José Martín Castro-Manzano[1], Axel Arturo Barceló-Aspeitia[1], and Alejandro Guerra-Hernández[2]

[1] Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México
Circuito Mario de la Cueva s/n Ciudad Universitaria, 04510, México, D.F., Mexico
jmcmanzano@hotmail.com, abarcelo@minerva.filosoficas.unam.mx
[2] Departamento de Inteligencia Artificial, Universidad Veracruzana,
Sebastián Camacho No. 5, 91000, Xalapa, Ver., Mexico
aguerra@uv.mx

Abstract. Defeasible logics have been mainly developed to reason about beliefs but have been barely used to reason about temporal structures; meanwhile, intentional logics have been mostly used to reason about intentional states and temporal behavior but most of them are monotonic. So, a defeasible temporal logic for intentional reasoning has not been developed yet. In this work we propose a defeasible temporal logic with the help of some temporal semantics and a non-monotonic framework in order to model intentional reasoning. We also show the consistency and soundness of the system.

Keywords: Defeasible logic, temporal logic, BDI logic.

1 Introduction

Intentional reasoning is a form of logical reasoning that uses beliefs and intentions during time. It has been mainly modeled via BDI logics, for instance [21,23,25]; however, there are two fundamental problems with such approaches: in first place, human reasoning is not and should not be monotonic [17], and thus, the logical models should be non-monotonic; and in second place, intentional states should respect temporal norms, and so, the logical models need to be temporal as well. Thus, the proof process of intentional reasoning has to have some sort of control over time and has to take into account a form of non-monotonic reasoning using beliefs and intentions.

In the state of the art defeasible logics have been mainly developed to reason about beliefs [19] but have been barely used to reason about temporal structures [11]; on the other hand, intentional logics have been mostly used to reason about intentional states and temporal behavior but most of them are monotonic. In order to solve the double problem mentioned above, our main contribution is the adaptation and extension for $CTL_{AgentSpeak(L)}$ [13] semantics with a non-monotonic framework. So, a defeasible temporal logic for intentional reasoning is proposed. We also show its consistency and soundness.

The relevance of this work becomes clear once we notice that, although intentions have received a lot of attention, their dynamic features have not been studied completely [24]. There are formal theories of intentional reasoning [6,14,21,23,25] but very few of them consider the revision of intentions [24] or the non-monotonicity of intentions [10] as legitimate research topics, which we find odd since the foundational theory guarantees that such research is legitimate and necessary [4]. Recent works confirm the status of this emerging area [10,24,18].

In Section 2 we discuss the case of intentional reasoning as a case of non-monotonic reasoning and we expose a non-monotonic framework for intentional reasoning. In Section 3 we display the system, its consistency and soundness. Finally, in Section 4 we discuss the results and we mention future work.

2 Non-monotonicity of Intentional Reasoning

The BDI models based upon Bratman's theory [4] tend to interpret intentions as a unique fragment [21,23,25] while Bratman's richer framework distinguished three classes of intentions: deliverative, non-deliverative and policy-based. In particular, policy-based intentions are of great importance given their structure and behavior: they have the form of rules and behave like plans. These remarks are relevant because the existing formalisms, despite of recognizing the intimate relationship between plans and intentions, seem to forget that intentions behave like plans.

As Bratman has argued, plans are intentions as well [4]. In this way we can set policy-based intentions to be structures $te : ctx \leftarrow body$ [2] (see Table 1). Now, consider the next example for sake of argument: $on(X, Y) \leftarrow put(X, Y)$. This intention tells us that, for an agent to achieve $on(a, b)$, it typically has to put a on b. If we imagine such an agent is immersed in a dynamic environment, of course the agent will try to put, typically, a on b; nevertheless, a *rational* agent would only do it as long as it is *possible*.

Thus, it results quite natural to talk about some intentions that are maintained typically but not absolutely. And so, it is reasonable to conclude that intentions, and particularly policy-based, allow defeasible intentional reasoning [10]. However, the current BDI models are monotonic and non-monotonic logics are barely used to reason about time [11] or intentional states. Thus, a defeasible temporal logic for intentional reasoning has not been developed yet. So, for example, standard First Order Logic is an instance of monotonic atemporal reasoning; default logic [22] is an instance of non-monotonic atemporal reasoning. In turn, BDI logic [21,23,25] is an example of temporal but monotonic reasoning. Our proposal is a case of temporal and non-monotonic reasoning.

Traditional BDI models formalize intentional reasoning in a monotonic way [6,14,21,23,25], while our proposal aims to do it non-monotonically. As a working example consider the following scenario under the traditional approach: an agent intends to acquire its PhD, $\mathsf{INT}(phd)$, and there is a rule $phd \Rightarrow exam$, then it follows that $\mathsf{INT}(exam)$. It is not hard to notice that this reasoning schema looks familiar. In knowledge bases it is known as the problem of logical omniscience [15]. Around intentions is it called the problem of collateral effect [4,16]. The schema above, $\mathsf{INT}(\phi) \wedge \phi \Rightarrow \psi \vdash \mathsf{INT}(\psi)$, is an example of collateral effect that does not allow us to distinguish between intentions that are maintained typically but not absolutely.

Despite great avances in this area, if we take into account the philosophical foundations of rational agency [4], it is not hard to see that most BDI logics fail to grasp all the properties of intentions: functional properties like proactivity, admissibility and inertia; descriptive properties like partiality, hierarchy and dynamism; and of course, the normative properties: internal consistency, strong consistency and means-end coherence. The explanation of these properties can be found in [4]. Following these ideas we propose the next framework:

Definition 1 *(Non-monotonic intentional framework) A non-monotonic intentional framework is a tuple* $\langle B, I, F_B, F_I, \vdash, \mid\!\sim, \dashv, \sim\!\mid, \succ \rangle$ *where:*

- B *denotes the belief base.*
- I *denotes the set of intentions.*
- $F_B \subseteq B$ *denotes the basic beliefs.*
- $F_I \subseteq I$ *denotes the basic intentions.*
- $\vdash$ *and* $\dashv$ *are strong consequence relations.*
- $\mid\!\sim$ *and* $\sim\!\mid$ *are weak consequence relations.*
- $\succ \subseteq I^2$ *s.t.* $\succ$ *is acyclic.*

With the help of this framework we can represent the non-monotonic nature of intentional reasoning. We assume a commitment strategy embedded in the agent architecture, i.e, we assume the inertia of intentions by a fixed mechanism that is single-minded [20], because if there is no commitment or the agent is blindly-committed, there is no sense in talking about inertia [12,13], i.e., in reconsidering intentions.

As usual, B denotes the beliefs base, which are literals. F_B stands for the beliefs that are considered basic; and similarly F_I stands for intentions considered as basic. Each intention $\phi \in I$ is a structure $te : ctx \leftarrow body$ where te represents the goal of the intention –so we preserve proactivity–, ctx a context and the rest denotes the body. When ctx or $body$ are empty we write $te : \top \leftarrow \top$ or just te.

We also preserve internal consistency by allowing the context of an intention, $ctx(\phi)$, $ctx(\phi) \in B$ and by letting te be the head of the intention. So, strong consistency is implied by internal consistency (given that strong consistency is $ctx(\phi) \in B$). Means-end coherence is implied by admissibility and the hierarchy of intentions is represented by the order relation, which we require to be acyclic in order to solve conflicts between intentions. Again, all these features can be found in [4]. And with this framework we can arrange a notion of inference where we say that ϕ is strongly (weakly) derivable from a sequence Δ iff there is a proof of $\Delta \vdash \phi$ ($\Delta \mid\!\sim \phi$). And also, that ϕ is not strongly (weakly) provable iff there is a proof of $\Delta \dashv \phi$ ($\Delta \sim\!\mid \phi$), where $\Delta = \langle B, I \rangle$.

3 Formal Model

In this work we adopt AgentSpeak(L) [21] because it has a well defined operational semantics. The problem, however, is that these particular semantics exclude modalities which are important to represent intentional states. To avoid this problem we use $CTL_{AgentSpeak(L)}$ [13] as a logical tool for the formal specification. Of course, initially, the approach is similar to a BDI^{CTL} system defined after $B^{KD45}D^{KD}I^{KD}$ with the temporal operators: *next* ($\bigcirc$), *eventually* ($\Diamond$), *always* ($\Box$), *until* (U), *optional* (E), *inevitable* (A), and so on, defined after $CTL*$ [7,9]. In this section we are going to expose the syntax and semantics of $CTL_{AgentSpeak(L)}$.

3.1 Syntax of AgentSpeak(L)

An agent ag is formed by a set of plans ps and beliefs bs (grounded literals). Each plan has the form $te : ctx \leftarrow h$. The context ctx of a plan is a literal or a conjunction of them. A non empty plan body h is a finite sequence of actions $A(t_1, \ldots, t_n)$, goals g (achieve ! or test ? an atomic formula $P(t_1, \ldots, t_n)$), or beliefs updates u (addition + or deletion −). $\top$ denotes empty elements, e.g., plan bodies, contexts, intentions. The trigger events te are updates (addition or deletion) of beliefs or goals. The syntax is shown in Table 1.

$ag ::= bs \quad ps$	$h ::= h_1; \top \mid \top$
$bs ::= b_1 \ldots b_n \ (n \geq 0)$	$h_1 ::= a \mid g \mid u \mid h_1; h_1$
$ps ::= p_1 \ldots p_n \ (n \geq 1)$	$at ::= P(t_1, \ldots, t_n) \ (n \geq 0)$
$p ::= te : ctx \leftarrow h$	$a ::= A(t_1, \ldots, t_n) \ (n \geq 0)$
$te ::= +at \mid -at \mid +g \mid -g$	$g ::= !at \mid ?at$
$ctx ::= ctx_1 \mid \top$	$u ::= +b \mid -b$
$ctx_1 ::= at \mid \neg at \mid ctx_1 \wedge ctx_1$	

Table 1. Sintax of AgentSpeak(L) adapted from [2].

3.2 Semantics of AgentSpeak(L)

The operational semantics of AgentSpeak(L) are defined by a transition system, as showed in Figure 1, 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 $\{i, i', \ldots, n\}$ s.t. $i \in I$ is a stack of partially instantiated plans $p \in ps$; E is a set of events $\{\langle te, i \rangle, \langle te', i' \rangle, \ldots, n\}$, 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.

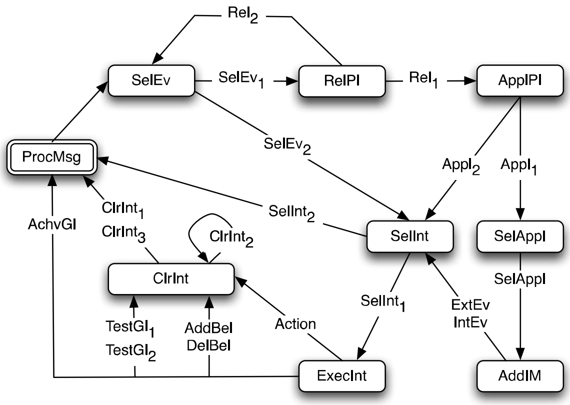


Fig. 1. The interpreter for AgentSpeak(L) as a transition system.

- 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).
- 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. $bs \models ctx$); ι, ϵ and ρ register, respectively, the intention, the event and the current plan during an agent execution.
- The label $s \in \{SelEv, RelPl, AppPl, SelAppl, SelInt, AddIM, ExecInt, ClrInt, ProcMsg\}$ indicates the current step in the reasoning cycle of the agent.

Under such semantics a run is a set $Run = \{(\sigma_i, \sigma_j) | \Gamma \vdash \sigma_i \rightarrow \sigma_j\}$ where Γ is the transition system defined by AgentSpeak(L) operational semantics and σ_i, σ_j are agent configurations.

3.3 Syntax of $BDI^{CTL}_{AS(L)}$

$CTL_{AgentSpeak(L)}$ may be seen as an instance of BDI^{CTL}. Similar approaches have been accomplished for other programming languages [8]. The idea is to define some BDI^{CTL} semantics in terms of AgentSpeak(L) structures. So, we need a language able to express temporal and intentional states. Thus, we require in first place some way to express these features.

Definition 2 *(Syntax of $BDI^{CTL}_{AS(L)}$) If ϕ is an AgentSpeak(L) atomic formula, then* $\mathsf{BEL}(\phi)$, $\mathsf{DES}(\phi)$ *and* $\mathsf{INT}(\phi)$ *are well formed formulas of* $BDI^{CTL}_{AS(L)}$.

To specify the temporal behavior we use $CTL*$ in the next way.

Definition 3 *($BDI^{CTL}_{AS(L)}$ temporal syntax) Every $BDI^{CTL}_{AS(L)}$ formula is a state formula s:*

- $s ::= \phi | s \wedge s | \neg s$
- $p ::= s | \neg p | p \wedge p | \mathsf{E}p | \mathsf{A}p | \bigcirc p | \Diamond p | \Box p | p \; \mathsf{U} \; p$

3.4 Semantics of $BDI^{CTL}_{AS(L)}$

Initially the semantics of BEL, DES and INT is adopted from [3]. So, we use the next function:

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

which gives us the set of atomic formulas (at) attached to an achievement goal ($+!$) and $i[p]$ denotes the stack of intentions with p at the top.

Definition 4 *($BDI^{CTL}_{AS(L)}$ semantics) The operators* BEL*,* DES *and* INT *are defined in terms of an agent ag and its configuration $\langle ag, C, M, T, s\rangle$:*

$$\mathsf{BEL}_{\langle ag,C,M,T,s\rangle}(\phi) \equiv \phi \in bs$$

$$\mathsf{INT}_{\langle ag,C,M,T,s\rangle}(\phi) \equiv \phi \in \bigcup_{i \in C_I} agoals(i) \vee \bigcup_{\langle te,i\rangle \in C_E} agoals(i)$$

$$\mathsf{DES}_{\langle ag,C,M,T,s\rangle}(\phi) \equiv \langle +!\phi, i\rangle \in C_E \vee \mathsf{INT}(\phi)$$

where C_I denotes current intentions and C_E suspended intentions.

We have a defeasible framework for intentions that lacks temporal representation; while the BDI temporal model described before grasps the temporal representation but lacks non-monotonicity. The next step is a system denoted by $NBDI$ because it has a non-monotonic behavior. An intention ϕ in $NBDI^{CTL}_{AS(L)}$ is a structure $\langle g : ctx \leftarrow body\rangle$ where g is the head, ctx is the context and $body$ is the body of the rule. We will denote an intention ϕ with head g by $\phi[g]$. Also, a negative intention is denoted by $\phi[g^c]$, i.e., the intention ϕ with $\neg g$ as the head.

The semantics of this theory requires a Kripke structure $K = \langle S, R, V\rangle$ where S is the set of agent configurations, R is an access relation defined after the transition system Γ and V is a valuation function that goes from agent configurations to true propositions in those states.

Definition 5 *Let $K = \langle S, \Gamma, V \rangle$, then:*

- *S is a set of agent configurations $c = \langle ag, C, M, T, s \rangle$.*
- *$\Gamma \subseteq S^2$ is a total relation s.t. for all $c \in \Gamma$ there is a $c' \in \Gamma$ s.t. $(c, c') \in \Gamma$.*
- *V is valuation s.t.:*
  - $V_{\mathsf{BEL}}(c, \phi) = \mathsf{BEL}_c(\phi)$ *where* $c = \langle ag, C, M, T, s \rangle$.
  - $V_{\mathsf{DES}}(c, \phi) = \mathsf{DES}_c(\phi)$ *where* $c = \langle ag, C, M, T, s \rangle$.
  - $V_{\mathsf{INT}}(c, \phi) = \mathsf{INT}_c(\phi)$ *where* $c = \langle ag, C, M, T, s \rangle$.
- *Paths are sequences of configurations $c_0, \ldots, c_n$ s.t. $\forall i (c_i, c_{i+1}) \in R$. We use x^i to indicate the i-th state of path x. Then:*

S1 $K, c \models \mathsf{BEL}(\phi) \Leftrightarrow \phi \in V_{\mathsf{BEL}}(c)$
S2 $K, c \models \mathsf{DES}(\phi) \Leftrightarrow \phi \in V_{\mathsf{DES}}(c)$
S3 $K, c \models \mathsf{INT}(\phi) \Leftrightarrow \phi \in V_{\mathsf{INT}}(c)$
S4 $K, c \models \mathsf{E}\phi \Leftrightarrow \exists x = c_1, \ldots \in K | K, x \models \phi$
S5 $K, c \models \mathsf{A}\phi \Leftrightarrow \forall x = c_1, \ldots \in K | K, x \models \phi$
P1 $K, c \models \phi \Leftrightarrow K, x^0 \models \phi$ *where ϕ is a state formula.*
P2 $K, c \models \bigcirc\phi \Leftrightarrow K, x^1 \models \phi$.
P3 $K, c \models \Diamond\phi \Leftrightarrow K, x^n \models \phi$ *for* $n \geq 0$
P4 $K, c \models \Box\phi \Leftrightarrow K, x^n \models \phi$ *for all* n
P5 $K, c \models \phi \mathsf{U} \psi \Leftrightarrow \exists k \geq 0$ *s.t.* $K, x^k \models \psi$ *and for all* j, k, $0 \leq j < k | K, c^j \models \phi$ *or* $\forall j \geq 0 : K, x^j \models \phi$

We have four cases of proof: if the sequence is $\Delta \vdash \phi$, we say ϕ is strongly provable; if it is $\Delta \dashv \phi$ we say ϕ is not strongly provable. If is $\Delta \mid\!\sim \phi$ we say ϕ is weakly provable and if it is $\Delta \sim\!\mid \phi$, then ϕ is not weakly provable.

Definition 6 *(Proof) A proof of ϕ from Δ is a finite sequence of beliefs and intentions satisfying:*

1. $\Delta \vdash \phi$ *iff*
   1.1. $\Box\mathsf{A}(\mathsf{INT}(\phi))$ *or*
   1.2. $\Box\mathsf{A}(\exists\phi[g] \in F_I : \mathsf{BEL}(ctx(\phi)) \wedge \forall\psi[g'] \in body(\phi) \vdash \psi[g'])$
2. $\Delta \mid\!\sim \phi$ *iff*
   2.1. $\Delta \vdash \phi$ *or*
   2.2. $\Delta \dashv \neg\phi$ *and*
      2.2.1. $\Diamond\mathsf{E}(\mathsf{INT}(\phi)\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi)))$ *or*
      2.2.2. $\Diamond\mathsf{E}(\exists\phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall\psi[g'] \in body(\phi) \mid\!\sim \psi[g'])$ *and*
         2.2.2.1. $\forall\gamma[g^c] \in I$, $\gamma[g^c]$ *fails at* Δ *or*
         2.2.2.2. $\psi[g'] \succ \gamma[g^c]$
3. $\Delta \dashv \phi$ *iff*
   3.1. $\Diamond\mathsf{E}(\mathsf{INT}(\neg\phi))$ *and*
   3.2. $\Diamond\mathsf{E}(\forall\phi[g] \in F_I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \dashv \psi)$
4. $\Delta \sim\!\mid \phi$ *iff*
   4.1. $\Delta \dashv \phi$ *and*
   4.2. $\Delta \vdash \neg\phi$ *or*
      4.2.1. $\Box\mathsf{A}\neg(\mathsf{INT}(\phi)\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi)))$ *and*
      4.2.2. $\Box\mathsf{A}(\forall\phi[g^c] \in I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \sim\!\mid \psi[g'])$ *or*
         4.2.2.1. $\exists\gamma[g^c] \in I$ *s.t.* $\gamma[g^c]$ *succeds at* Δ *and*
         4.2.2.2. $\psi[g'] \not\succ \gamma[g^c]$

3.5 Consistency

The next statements are quite straightforward.

Proposition 1 *(Subalterns$_1$) If $\vdash \phi$ then $\mid\!\sim \phi$.*

Proof. Let us assume that $\vdash \phi$ but not $\mid\!\sim \phi$, i.e., $\sim\!\mid \phi$. Then, given $\vdash \phi$ we have two general cases. Case 1: given the initial assumption that $\vdash \phi$, by Definition 6 item 1.1, we have that $\Box\mathsf{A}(\mathsf{INT}(\phi))$. Now, given the second assumption, i.e., that $\sim\!\mid \phi$, by Definition 6 item 4.1, we have $\dashv \phi$. And so, $\Diamond\mathsf{E}(\mathsf{INT}(\neg\phi))$, and thus, by the temporal semantics, we get $\neg\phi$; however, given the initial assumption, we also obtain ϕ, which is a contradiction.

Case 2: given the assumption that $\vdash \phi$, by Definition 6 item 1.2, we have that $\exists\phi[g] \in F_I : \mathsf{BEL}(ctx(\phi)) \wedge \forall\psi[g'] \in body(\phi) \vdash \psi[g']$. Now, given the second assumption, that $\sim\!\mid \phi$, we also have $\dashv \phi$ and so we obtain $\Diamond\mathsf{E}(\forall\phi[g] \in F_I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \dashv \psi)$, and thus we can obtain $\forall\phi[g] \in F_I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \dashv \psi)$ which is $\neg(\exists\phi[g] \in F_I : \mathsf{BEL}(ctx(\phi)) \wedge \forall\psi[g'] \in body(\phi) \vdash \psi[g'])$. ■

Corollary 1 *(Subalterns$_2$) If $\sim\!\mid \phi$ then $\dashv \phi$.*

Proposition 2 *(Contradictories$_1$) There is no ϕ s.t. $\vdash \phi$ and $\dashv \phi$.*

Proof. Assume that there is a ϕ s.t. $\vdash \phi$ and $\dashv \phi$. If $\dashv \phi$ then, by Definition 6 item 3.1, $\Diamond\mathsf{E}(\mathsf{INT}(\neg\phi))$. Thus, by proper semantics, we can obtain $\neg\phi$. However, given that $\vdash \phi$ it also follows that ϕ, which is a contradiction. ■

Corollary 2 *(Contradictories$_2$) There is no ϕ s.t. $\mid\!\sim \phi$ and $\sim\!\mid \phi$.*

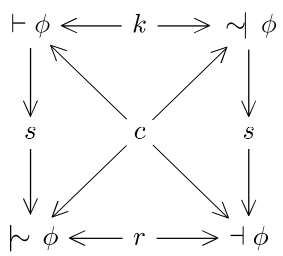
Proposition 3 *(Contraries) There is no ϕ s.t. $\vdash \phi$ and $\sim\!\mid \phi$.*

Proof. Assume there is a ϕ such that $\vdash \phi$ and $\sim\!\mid \phi$. By Proposition 1, it follows that $\mid\!\sim \phi$, but that contradicts the assumption that $\sim\!\mid \phi$ by Corollary 2. ■

Proposition 4 *(Subcontraries) For all ϕ either $\mid\!\sim \phi$ or $\dashv \phi$.*

Proof. Assume it is not the case that for all ϕ either $\mid\!\sim \phi$ or $\dashv \phi$. Then there is ϕ s.t. $\sim\!\mid \phi$ and $\vdash \phi$. Taking $\sim\!\mid \phi$ it follows from Corollary 1 that $\dashv \phi$. By Proposition 2 we get a contradiction with $\vdash \phi$. ■

Considering these results, we get the next square of opposition where c denotes contradictories, s subalterns, k contraries and r subcontraries.



These results represent the following properties: Proposition 1 and Corollary 1 represent supraclassicality; Proposition 2 and Corollary 2 stand for consistency while the remaining statements specify the coherence of the square, and thus, the overall coherence of the system.

If we recover our working example, the scenario in which an agent intends to acquire its PhD, and we set the next configuration Δ of beliefs and intentions: $F_B = \{\top\}$, $B = \{scolarship\}$, $F_I = \{research : \top \leftarrow \top\}$, $I = \{phd : \top \leftarrow thesis, exam; thesis : scolarship \leftarrow research; exam : \top \leftarrow research\}$. And suppose we send the query: *phd?* The search of intentions with head *phd* in F_I fails, thus the alternative $\vdash \phi[phd]$ does not hold. Thus, we can infer, by contradiction rule (Proposition 2), that it is not strongly provable that *phd*, i.e., that eventually in some state the intention *phd* does not hold. Thus, the result of the query should be that the agent will get its PhD defeasibly under the Δ configuration. On the contrary, the query *research?* will succedd as $\vdash \phi[research]$, and thus, we would say *research* is both strongly and weakly provable (Proposition 1).

3.6 Soundness

The idea is to show the framework is sound with respect to its semantics. Thus, as usual, we will need some notions of satisfaction and validity.

Definition 7 *(Satisfaction) A formula ϕ is true in K iff ϕ is true in all configurations σ in K. This is to say, $K \models \phi \Leftrightarrow K, \sigma \models \phi$ for all $\sigma \in S$.*

Definition 8 *(Run of an agent in a model) Given an initial configuration β, a transition system Γ and a valuation V, $K^\beta_\Gamma = \left\langle S^\beta_\Gamma, R^\beta_\Gamma, V \right\rangle$ denotes a run of an agent in a model.*

Definition 9 *(Validity) A formula $\phi \in BDI^{CTL}_{AS(L)}$ is true for any agent run in Γ iff $\forall K^\beta_\Gamma \models \phi$*

Further, we will denote $(\exists K^\beta_\Gamma \models \phi \ \mathsf{U} \ \neg \mathsf{BEL}(ctx(\phi))) \vee \models \phi$ by $\approx\!\!\!| \ \phi$. We can observe, moreover, that $\models \phi \geq \approx\!\!\!| \ \phi$ and $\approx\!\!\!\dashv \ \phi \geq =\!\!\!| \ \phi$. With these remarks we should find a series of *translations* s.t.: $\vdash \phi \longrightarrow \forall K^\beta_\Gamma \models \phi \longrightarrow \models \phi$ and $\models \phi \longrightarrow \approx\!\!\!| \ \phi$.

Proposition 5 *If $\vdash \phi$ then $\models \phi$.*

Proof. Base case. Taking Δ_i as a sequence with $i = 1$. If we assume $\vdash \phi$, we have two subcases. First subcase is given by Definition 6 item 1.1. Thus we have $\Box\mathsf{A}(\mathsf{INT}(\phi))$. This means, by Definition 5 items P4 and S5 and Definition 4, that for all paths and all states $\phi \in C_I \vee C_E$. We can represent this expression, by way of a translation, in terms of runs. Since paths and states are sequences of agent configurations we have that $\forall K^\beta_\Gamma \models \phi$, which implies $\models \phi$. Second subcase is given by Definition 6 item 1.2, which in terms of runs means that for all runs $\exists \phi[g] \in F_I : \mathsf{BEL}(ctx(\phi)) \wedge \forall \psi[g'] \in body(\phi) \vdash \psi[g']$. Since Δ_1 is a single step,

$body(\phi) = \top$ and for all runs $\mathsf{BEL}(ctx(\phi)))$, $ctx(\phi) \in F_B$. Then $\forall K_\Gamma^\beta \models \phi$ which, same as above, implies $\models \phi$.

Inductive case. Let us assume that for $n \leq k$, if $\Delta_n \vdash \phi$ then $\Delta \models \phi$. And suppose Δ_{n+1}. Further, suppose $\Delta_n \vdash \phi$, then we have two alternatives. First one being, by Definition 6 item 1.1, that we have an intention ϕ s.t. $ctx(\phi) = body(\phi) = \top$. Since $body(\phi)$ is empty, it trivially holds at n, and by the induction hypothesis, $body(\phi) \subseteq \Delta_{n+1}$, and thus $\models \phi$. Secondly, by Definition 6 item 1.2, for all runs $\exists \phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall \psi[g'] \in body(\phi) \vdash \psi[g']$. Thus, for all runs $n, \forall \psi[g'] \in body(\phi) \vdash \psi[g']$, and so by the induction hypothesis, $body(\phi) \subseteq \Delta_{n+1}$, i.e., $\Delta \vdash \psi[g']$. Therefore, $\models \phi$. ■

Proposition 6 *If* $\mid\!\sim \phi$ *then* $\models\!\!\approx \phi$.

Base case. Taking Δ_i as a sequence with $i = 1$. Let us suppose $\mid\!\sim \phi$. Then we have two subcases. The first one is given by Definition 6 item 2.1. So, we have that $\vdash \phi$ which, as we showed above, already implies $\models \phi$. On the other hand, by item 2.2, we have $\dashv \neg\phi$ and two alternatives. The first alternative, item 2.2.1, is $\Diamond \mathsf{E}(\mathsf{INT}(\phi)\ \mathsf{U}\ \neg \mathsf{BEL}(ctx(\phi)))$. Thus, we can reduce this expression by way of Definition 5 items P3 and S4, to a translation in terms of runs: $\exists K_\Gamma^\beta \models \phi\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi))$, which implies $\models\!\!\approx \phi$. The second alternative comes from item 2.2.2, $\Diamond \mathsf{E}(\exists \phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall \psi[g'] \in body(\phi) \mid\!\sim \psi[g'])$ which in terms of runs means that for some run $\exists \phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall \psi[g'] \in body(\phi) \mid\!\sim \psi[g']$, but Δ_1 is a single step, and thus $body(\phi) = \top$. Thus, there is a run in which $\exists \phi[g] \in I : \mathsf{BEL}(ctx(\phi))$, i.e., $(\exists K_\Gamma^\beta \models (\phi\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi)))$ by using the weak case of Definition 6 P5. Thus, by addition, $(\exists K_\Gamma^\beta \models (\phi\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi))) \vee \models \phi$, and therefore, $\models\!\!\approx \phi$.

Inductive case. Let us assume that for $n \leq k$, if $\Delta_n \mid\!\sim \phi$ then $\Delta \models\!\!\approx \phi$. And suppose Δ_{n+1}. Assume $\Delta_n \mid\!\sim \phi$. We have two alternatives. The first one is given by Definition 6 item 2.1, i.e., $\vdash \phi$, which already implies $\models \phi$. The second alternative is given by item 2.2, $\Delta \dashv \neg\phi$ and two subcases: $\Diamond \mathsf{E}(\mathsf{INT}(\phi)\ \mathsf{U}\ \neg\mathsf{BEL}(ctx(\phi)))$ or $\Diamond \mathsf{E}(\exists \phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall \psi[g'] \in body(\phi) \mid\!\sim \psi[g'])$. If we consider the first subcase there are runs n which comply with the definition of $\models\!\!\approx \phi$. In the remaining subcase we have $\forall \psi[g'] \in body(\phi) \mid\!\sim \psi[g']$, since $body(\phi) \subseteq \Delta_n$, by the induction hypothesis $\Delta \mid\!\sim \psi[g']$, and thus, $\Delta_{n+1} \mid\!\sim \phi$, i.e., $\models\!\!\approx \phi$. ■

Also, we can find a series of translations for the remaining fragments:

Corollary 3 *If* $\dashv \phi$ *then* $=\!\!\mid \phi$*; and if* $\sim\!\!\mid \phi$ *then* $\approx\!\!\mid \phi$

4 Conclusion

The formal model described above attempts to represent temporal and non-monotonic features of intentional reasoning. We observed the model preserves supraclassicality, consistency and soundness.

Currently we are trying to find relations between the notion of inference of this system and a notion of intention revision [5]. For example, is it the case that intentions strongly proved cannot be contracted? And is it the case that intentions weakly proved can be revised? Conversely, revised intentions are defeasible? And so on. Plus, since the model of revision is related to AgentSpeak(L), we foresee implementations that may follow organically.

Acknowledgements. The authors would like to thank the anonymous reviewers for all the useful comments and precise corrections. First author is supported by the CONACyT scholarship 214783.

References

1. Alchourrón, C. E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic, 50, 510–530 (1985)
2. Bordini, R.H., Wooldridge, M., Hübner, J.F.: Programming Multi-Agent Systems in AgentSpeak using Jason (Wiley Series in Agent Technology). John Wiley & Sons (2007)
3. Bordini, R.H., Moreira, Á.F.: Proving BDI Properties of Agent-Oriented Programming Languages. Annals of Mathematics and Artificial Intelligence, 42, 197–226 (2004)
4. Bratman, M.E.: Intention, Plans, and Practical Reason. Cambridge University Press (1999)
5. Castro-Manzano, J.M., Barceló-Aspeitia, A.A., Guerra-Hernández, A.: Intentional learning procedures as intention revision mechanisms. Mexican International Conference on Artificial Intelligence, 51–56 (2010)
6. Cohen, P., Levesque, H.: Intention is choice with commitment. Artificial Intelligence, 42(3), 213–261 (1990)
7. Clarke, E.M. Jr., Grumberg, O.,Peled. D.A.: Model Checking. MIT Press (1999)
8. Dastani, M., van Riemsdijk, M.B., Meyer, J.C.: A grounded specification language for agent programs. Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems '07 AAMAS '07, 1–8 (2007)
9. Emerson, A.: Temporal and modal logic. Handbook of Theoretical Computer Science, Elsevier Science Publishers, 995–1072 (1995)
10. Governatori, G., Padmanabhan, V. and Sattar, A.: A Defeasible Logic of Policy-based Intentions. Proceedings of the 15th Australian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence. LNAI-2557, Springer Verlag (2002)
11. Governatori, G., Terenziani, P.: Temporal Extensions to Defeasible Logic. Proceedings of the 20th Australian joint conference on Advances in artificial intelligence, 476–485 (2007)
12. Guerra-Hernández, A., Castro-Manzano, J.M., El-Fallah-Seghrouchni, A.: Toward an AgentSpeak(L) Theory of Commitment and Intentional Learning. Proceedings of the 7th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence MICAI 2008, LNCS, vol. 5317, 848–858, Springer-Verlag, Berlin Heidelberg, (2008)
13. Guerra-Hernández, A., Castro-Manzano, J.M., El-Fallah-Seghrouchni, A.: CTL AgentSpeak(L): a Specification Language for Agent Programs. J. Algorithms, 64(1), 31–40 (2009)

14. Konolige, K., Pollack, M. E.: A representationalist theory of intentions. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-93), 390–395 (1993)
15. K. M. Sim: Epistemic Logic and Logical Omniscience: A Survey. International Journal of Intelligent Systems, vol. 12, 57–81, John Wiley and Sons Inc. (1997)
16. Linder, B. Van: Modal Logic for Rational Agents. PhD thesis, Department of Computer Science, Utrecht University, 19th June (1996)
17. Nute, D.: Defeasible logic. INAP 2001, LNAI 2543M, Springer-Verlag, 151–169 (2003)
18. Icard, Th., Pacuit. E., Shoham, Y.: Joint revision of belief and intention. Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning, (2010)
19. Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. In D. Gabbay and F. Guenthner (eds.), Handbook of Philosophical Logic, second edition, Vol 4, Kluwer Academic Publishers, 219–318 (2002)
20. Rao, A.S., Georgeff, M.P.: Modelling Rational Agents within a BDI Architecture. In: Huhns, M.N., Singh, M.P., (eds.) Readings in Agents, Morgan Kaufmann, 42–55 (1998)
21. 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, Springer, Heidelberg, 42–55 (1996)
22. Reiter, R.: A logic for default reasoning. Artificial Intelligence, 13, 81–132, (1980)
23. 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, MIT Press, Cambridge, 331–376 (1999)
24. van der Hoek, W., Jamroga, W., Wooldridge, M.: Towards a theory of intention revision. Synthese, 155(2), 265–290 (2007)
25. Wooldridge, M.: Reasoning about Rational Agents. MIT Press (2000).