Modelling Intentional Reasoning with Defeasible and Temporal Logic

José Martín Castro-Manzano

Escuela de Filosofía
Universidad Popular Autónoma del Estado de Puebla
21 sur 1103 Barrio de Santiago, Puebla, México 72410
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
Circuito Mario de la Cueva s/n Ciudad Universitaria, México, D.F., México 04510
josemartin.castro@upaep.mx

Abstract. We follow the hypothesis that intentional reasoning is a form of logical reasoning *sui generis* by its double nature: temporal and defeasible. Then we briefly describe a formal framework that deals with these topics and we study the metalogical properties of its notion of inference. The idea is that intentional reasoning can be represented in a well-behaved defeasible logic and has the right to be called logical reasoning since it behaves, *mutatis mutandis*, as a logic, strictly speaking, as a non-monotonic logic.

Keywords: Defeasible logic, temporal logic, BDI logic, intention.

1 Introduction

The relationship between philosophy and computer science is very profound and unique [23]. Not only because these disciplines share some common historical data –like Leibniz's *mathesis universalis* [8]– and interesting anecdotes –like the correspondence between Newell and Russell [11]–, but more importantly because from the constant dialog that occurs within these disciplines we gain useful hypothesis, formal methods and functional analysis that may shed some light about different aspects of the nature of human behavior, specially under a cognitive schema. The cognitive schema we follow is the BDI model (that stands for *Beliefs*, *Desires* and *Intentions*) as originally exposed by Bratman [4] and formally developed by Rao and company [21,22]. The general aspect we study is the case of the non-monotonicity of intentional reasoning.

There is no doubt that reasoning using beliefs and intentions during time is a very common task, done on a daily basis; but the nature and the status of such kind of reasoning, which we will be calling intentional, are far from being clear and distinct. However, it would be blatantly false to declare that this study is entirely new, for there are recent efforts to capture some of these ideas already [13,16,19]. But, in particular, we can observe, on one side, the case of BDI logics [22,24] in order to capture and understand the nature of intentional

reasoning; and on the other side, the case of defeasible logics [20] to try to catch the status of non-monotonic reasoning.

The problem with these approaches, nevertheless, is that, in first place, human reasoning is not and should not be monotonic [18], and thus, the logical models should be non-monotonic, but the BDI techniques are monotonic; and in second place, intentional states should respect temporal norms, and so, the logical models need to be temporal as well, but the non-monotonic procedures do not consider the temporal or intentional aspect. So, in the state of the art, defeasible logics have been mainly developed to reason about beliefs [20] but have been barely used to reason about temporal structures [14]; on the other hand, intentional logics have been mostly used to reason about intentional states and temporal behavior but most of them are monotonic [7,21,24].

Under this situation our main contribution is a brief study of the nature and status of intentional reasoning following the hypothesis that intentional reasoning is a form of logical reasoning *sui generis* by its temporal and defeasible nature and we suggest that intentional reasoning has the right to be called *logical* since it behaves, *mutatis mutandis*, as a logic. In particular, this study is important by its own sake because defeasible reasoning has certain patterns of inference and therefore the usual challenge is to provide a reasonable description of these patterns. Briefly, the idea is that if monotony is not a property of intentional reasoning and we want to give an adequate description of its notion of inference, then we must study the metalogical properties of intentional inference that occur instead of monotony. Because once monotonicity is given up, a very organic question about the status of this kind of reasoning emerges: why should we consider intentional reasoning as an instance of a logic *bona fide*?

This paper is organized in the next way. In Section 2 we briefly expose what is understood as intentional reasoning. In Section 3 is our main contribution and finally, in Section 4 we sum up the results obtained.

2 Intentional reasoning

Two general requirements to be checked out while developing a logical framework are material and formal adequacy [1]. Material adequacy is about capturing an objective phenomenon. Formal adequacy has to do with the metalogical properties that a notion of logical consequence satisfies. The nature of intentional reasoning is related to a material aspect, while its status is directly connected with a formal one. During this study, due to reasons of space, we will focus mainly on the latter in order to argue that intentional reasoning can be modelled in a well-behaved defeasible logic, given that a well-behaved defeasible logic has to satisfy conditions of Supraclassicality, Reflexivity, Cut and Cautious Monotony [12].

But just to give some pointers about material adequacy, let us consider the next example for sake of explanation: assume there is an agent that has an intention of the form $on(X, Y) \leftarrow put(X, Y)$. This means that, for such 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, typ-

ically, a on b; nevertheless, a *rational* agent would only do it as long as it is *possible*; otherwise, we would say the agent is not rational. Therefore, it results quite natural to talk about some intentions that are maintained typically but not absolutely if we want to guarantee some level of rationality. And so, it is reasonable to conclude that intentions –in particular policy-based intentions [4]–, allow some form of defeasible reasoning [13] that must comply with some metalogical properties. But before we explore such properties, let us review some previous details.

The current logical systems that are used to model intentional reasoning are built in terms of what we call a bratmanian model. A bratmanian model is a model that *i)* follows general guidelines of Bratman's theory of practical reasoning [4], *ii)* uses the BDI architecture [21] to represent data structures and *iii)* configures notions of logical consequence based on relations between intentional states. There are several logics based upon bratmanian models, but we consider there are, at least, two important problems with the usual logics [7,22,24].

For one, such logics tend to interpret intentions as a unique fragment –usually represented by an opertator INT–, while Bratman's original theory distinguished three classes of intentions: deliberative, non-deliberative and policy-based. In particular, policy-based intentions are of great importance given their structure and behavior: they have the structure of complex rules and behave like plans. This remark is important for two reasons: because the existing formalisms, despite of recognizing the intimate relationship between plans and intentions, seem to forget that intentions behave like plans; and because the rule-like structure allows us to build a more detailed picture of the nature of intentional reasoning.

But the bigger problem is that these systems do not quite recognize that intentional reasoning has a temporal and defeasible nature. Intuitively, the idea is that intentional reasoning is temporal because intentions and beliefs are dynamic data structures, i.e., they change during time; but is also defeasible, because if these data structures are dynamic, their consequences may change. The bratmanian model we propose tries to respect this double nature by following the general guidelines of Bratman's theory of practical reasoning [4], so we distinguish functional (pro-activity, inertia, admissibility), descriptive (partiality, dynamism, hierarchy) and normative (internal, external consistency and coherence) properties that configure a notion of inference. To capture this notion of inference in a formal fashion the next framework is proposed in terms of $AgentSpeak(L)$[3] (see Appendix):

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.*

The item B denotes the beliefs, which are literals. F_B stands for the beliefs that are considered as 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. Also it is assumed that plans are *partially* instantiated.

Internal consistency is preserved by allowing the context of an intention denoted by $ctx(\phi)$, $ctx(\phi) \in B$ 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* will be implied by *admissibility* –the constraint that an agent will not consider contradictory options– and the *hierarchy* of intentions is represented by the order relation, which we require to be acyclic in order to solve conflicts between intentions. And with this framework we can arrange a notion of inference where we will say that ϕ is strongly (weakly) derivable from a sequence Δ if and only if there is a proof of $\Delta \vdash \phi$ ($\Delta \mid\sim \phi$). And also, that ϕ is not strongly (weakly) provable if and only if there is a proof of $\Delta \dashv \phi$ ($\Delta \sim\!\!\mid \phi$), where $\Delta = \langle B, I \rangle$.

2.1 The system $NBDI^{CTL}_{AS(L)}$

We start with $CTL_{AgentSpeak(L)}$ [15] as a logical tool for the formal specification (similar approaches have been accomplished for other programming languages [9]). 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 $CTL*$ [6,10].

Syntax of $BDI^{CTL}_{AS(L)}$ $CTL_{AgentSpeak(L)}$ may be seen as an instance of BDI^{CTL}. 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:*

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

Semantics of $BDI^{CTL}_{AS(L)}$ Initially the semantics of BEL, DES and INT is adopted from [2]. So, we assume the next function:

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

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.

And now some notation: 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 will require 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 such that 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 $\forall j \geq 0 : K, x^j \models \phi$*

A notion of inference comes in four cases: 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 $\Delta \mathrel{|\!\sim} \phi$ we say ϕ is weakly provable and if it is $\Delta \sim\!\!\!| \ \phi$, 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. $\Delta \mathrel{|\!\sim} \phi$ *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. $\Diamond\mathsf{E}(\exists\phi[g] \in I : \mathsf{BEL}(ctx(\phi)) \wedge \forall\psi[g'] \in body(\phi) \mathrel{|\!\sim} \psi[g'])$ *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. $\Diamond\mathsf{E}(\forall\phi[g] \in F_I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \dashv \psi[g'])$
4. $\Delta \sim\!\!\!| \ \phi$ *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. $\Box\mathsf{A}(\forall\phi[g] \in I : \neg\mathsf{BEL}(ctx(\phi)) \vee \exists\psi[g'] \in body(\phi) \sim\!\!\!| \ \psi[g'])$ *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 Formal adequacy

Once monotonicity is given up a very intuitive question arises: why should we consider intentional reasoning as an instance of a logic *bona fide*? We indirectly answer this question by arguing that intentional reasoning under this bratmanian model has some good properties.

3.1 Consistency

We suggest a square of opposition in order to depict logical relationships of consistency and coherence.

Proposition 1 *(Subalterns$_1$) If* $\vdash \phi$ *then* $\mathrel{|\!\sim} \phi$.

Proof. Let us assume that $\vdash \phi$ but not $\mathrel{|\!\sim} \phi$, i.e., $\sim\!\!\!| \ \phi$. 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 $\sim\!\!\!| \ \phi$, 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 $\sim\!\!\dashv \phi$, 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 $\sim\!\!\dashv \phi$ 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 $\sim\!\!\dashv \phi$.*

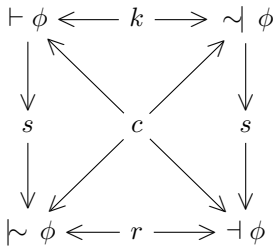
Proposition 3 *(Contraries) There is no ϕ s.t. $\vdash \phi$ and $\sim\!\!\dashv \phi$.*

Proof. Assume there is a ϕ such that $\vdash \phi$ and $\sim\!\!\dashv \phi$. By Proposition 1, it follows that $\mid\!\sim \phi$, but that contradicts the assumption that $\sim\!\!\dashv \phi$ 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. $\sim\!\!\dashv \phi$ and $\vdash \phi$. Taking $\sim\!\!\dashv \phi$ it follows from Corollary 1 that $\dashv \phi$. By Proposition 2 we get a contradiction with $\vdash \phi$. ■

These propositions form the next square of opposition where c denotes contradictories, s subalterns, k contraries and r subcontraries.



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.

Consider, for example, a 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).

3.2 Soundness

The framework is Sound with respect to its semantics.

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 $\forall K^{\beta}_{\Gamma} \models \phi$*

By denoting $(\exists K^{\beta}_{\Gamma} \models \phi \ \mathsf{U} \ \neg\mathsf{BEL}(ctx(\phi))) \vee \models \phi$ as $\models\!\!\approx \phi$, and assuming $\models \phi \geq \models\!\!\approx \phi$ and $\approx\!\!\models \phi \geq =\!\!\models \phi$, a series of *translations* can be found s.t.:

$$\vdash \phi \longrightarrow \forall K^{\beta}_{\Gamma} \models \phi \longrightarrow \models \phi$$
$$\searrow \qquad\qquad \downarrow$$
$$\vdash\sim \phi \longrightarrow \models\!\!\approx \phi$$

And also for the rest of the fragments:

$$\sim\!\!\dashv \phi \longrightarrow \exists K^{\beta}_{\Gamma} \models \neg\phi \wedge \forall K^{\beta}_{\Gamma} \models \neg(\phi \ \mathsf{U} \ \neg\mathsf{BEL}(ctx(\phi))) \longrightarrow \approx\!\!\dashv \phi$$
$$\searrow \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \downarrow$$
$$\dashv \phi \longrightarrow\!\!\dashv\!\!\approx \phi$$

Proposition 5 *The following relations hold:*

$$a)\ If \vdash \phi\ then \models \phi \qquad b)\ If \vdash\sim \phi\ then \models\!\!\approx \phi$$

Proof. Base case. Taking Δ_i as a sequence with $i = 1$.

Case a) 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 $\forall K_\Gamma^\beta \models \phi$, 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 $\models \phi$.

Case b) 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 $\mid\approx \phi$. 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, $\mid\approx \phi$.

Inductive case. Case a) 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, $\models \phi$.

Case b) Let us assume that for $n \leq k$, if $\Delta_n \mid\sim \phi$ then $\Delta \mid\approx \phi$. 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 $\mid\approx \phi$. 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., $\mid\approx \phi$. ■

Corollary 3 *The following relations hold:*

$$a)\ If\ \dashv \phi\ then\ =\!\mid\ \phi \qquad b)\ If\ \sim\!\mid\ \phi\ then\ \approx\!\mid\ \phi$$

3.3 Other formal properties

But there are other formal properties that may be used to explore and define the rationality of intentional reasoning, i.e., its good behavior. In first place, it results quite reasonable to impose Reflexivity on the consequence relation so that if $\phi \in \Delta$, then $\Delta \mid\sim \phi$.

Further, another reasonable property should be one that dictates that strong intentions imply weak intentions. In more specific terms, that if an intention ϕ follows from Δ in a monotonic way, then it must also follow according to a non-monotonic approach. Thus, in second place, we need the reasonable requirement that intentions strongly mantained have to be also weakly mantained, but no the other way around:

Proposition 6 *(Supraclassicality) If* $\Delta \vdash \phi$*, then* $\Delta \mid\sim \phi$*.*

Proof. See Proposition 1. ■

Another property, a very strong one, is Consistency Preservation. This property tells us that if some intentional set is classically consistent, then so is the set of defeasible consequences of it.

Proposition 7 *(Consistency preservation) If* $\Delta \mid\sim \bot$*, then* $\Delta \vdash \bot$*.*

Proof. Let us consider the form of the intention $\bot$. Such intention is the intention of the form $\phi \wedge \neg\phi$, which is, therefore, impossible to achieve, that is to say, for all agent runs, $\mid\sim \bot$ is never achieved. Thus $\Delta \mid\sim \bot$ is false, which makes the whole implication true. ■

And, if an intention ϕ is a consequence of Δ, then ψ is a consequence of Δ and ϕ only if it is already a consequence of Δ, because adding to Δ some intentions that are already a consequence of Δ does not lead to any *increase* of information. In terms of the size of a proof [1], such size does not affect the degree to which the initial information supports the conclusion:

Proposition 8 *(Cautious cut) If* $\Delta \mid\sim \phi$ *and* $\Delta, \phi \mid\sim \psi$ *then* $\Delta \mid\sim \psi$*.*

Proof. Let us start by transforming the original proposition into the next one: if $\Delta \nmid\sim \psi$ then it is not the case that $\Delta \mid\sim \phi$ and $\Delta, \phi \mid\sim \psi$. Further, this proposition can be transformed again: if $\Delta \nmid\sim \psi$ then either $\Delta \nmid\sim \phi$ or $\Delta, \phi \nmid\sim \psi$ from which, using Corollary 1, we can infer: if $\Delta \dashv \psi$ then either $\Delta \dashv \phi$ or $\Delta, \phi \dashv \psi$. Now, let us assume that $\Delta \dashv \psi$ but it is not the case that either $\Delta \dashv \phi$ or $\Delta, \phi \dashv \psi$, i.e., that $\Delta \dashv \psi$ but $\Delta \vdash \phi$ and $\Delta, \phi \vdash \psi$. Considering the expression $\Delta, \phi \vdash \psi$ we have two alternatives: either $\psi \in \mathit{body}(\phi)$ or $\psi \notin \mathit{body}(\phi)$. In the first case, given that $\Delta \vdash \phi$ then, since $\psi \in \mathit{body}(\phi)$ it follows that $\vdash \psi$, but that contradicts the assumption that $\Delta \dashv \psi$. In the remaining case, if $\Delta, \phi \vdash \psi$ but $\psi \notin \mathit{body}(\phi)$, then $\Delta \vdash \psi$, which contradicts the assumption that $\Delta \dashv \psi$. ■

If we go a little bit further, we should look for some form of Cautious Monotony as the converse of Cut in such a way that if ϕ is taken back into Δ that does not lead to any *decrease* of information, that is to say, that adding implicit information is a monotonic task:

Proposition 9 *(Cautious monotony) If* $\Delta \mid\sim \psi$ *and* $\Delta \mid\sim \gamma$ *then* $\Delta, \psi \mid\sim \gamma$*.*

Proof. Let us transform the original proposition: if $\Delta, \psi \nmid\sim \gamma$ then it is not the case that $\Delta \mid\sim \psi$ and $\Delta \mid\sim \gamma$. Thus, if $\Delta, \psi \nmid\sim \gamma$ then either $\Delta \nmid\sim \psi$ or

$\Delta \nsim \gamma$, and by Corollary 1, if $\Delta, \psi \dashv \gamma$ then either $\Delta \dashv \psi$ or $\Delta \dashv \gamma$. Now, let us suppose that $\Delta, \psi \dashv \gamma$ but it is false that either $\Delta \dashv \psi$ or $\Delta \dashv \gamma$, this is to say, that $\Delta, \psi \dashv \gamma$ and $\Delta \vdash \psi$ and $\Delta \vdash \gamma$. Regarding the expression $\Delta, \psi \dashv \gamma$ we have two alternatives: either $\gamma \in body(\psi)$ or $\gamma \notin body(\psi)$. In the first case, since $\gamma \in body(\psi)$ and $\Delta \dashv \psi$, then $\dashv \gamma$, which contradicts the assumption that $\Delta \vdash \gamma$. On the other hand, if we consider the second alternative, $\Delta \dashv \gamma$, but that contradicts the assumption that $\Delta \vdash \gamma$. ■

4 Conclusion

It seems reasonable to conclude that this bratmanian model of intentional reasoning captures relevant features of the nature of intentional reasoning and can be modelled in a well-behaved defeasible logic that clarifies its status, since it satisfies conditions of Consistency, Soundness, Supraclassicality, Reflexivity, Consistency Preservation, Cautious Cut and Cautious Monotony. In other words, it is plausible to conclude that intentional reasoning has the right to be called *logical reasoning* since it behaves, *mutatis mutandis*, as a logic, strictly speaking, as a non-monotonic logic.

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 [16]. There are formal theories of intentional reasoning [7,17,22,24] but very few of them consider the revision of intentions [16] or the non-monotonicity of intentions [13] 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 [13,16,19].

Acknowledgements. The author would like to thank the anonymous reviewers for their helpful comments and precise corrections; and the School of Philosophy at UPAEP for all the assistance. This work has also been supported by the CONACyT scholarship 214783.

References

1. Antonelli, A.: Grounded Consequence for Defeasible Logic. Cambridge: Cambridge University Press (2005)
2. Bordini, R.H., Moreira, Á.F.: Proving BDI properties of agent-oriented programming languages. Annals of Mathematics and Artificial Intelligence 42, 197–226 (2004)
3. Bordini, R.H., Hübner, J.F., Wooldridge, M.: Programming Multi-Agent Systems in AgentSpeak using Jason. Wiley, England (2007)
4. Bratman, M.: Intention, Plans, and Practical Reason. Harvard University Press, Cambridge (1987)
5. Castro-Manzano, J.M., Barceló-Aspeitia. A.A. and Guerra-Hernández, A.: Consistency and soundness for a defeasible logic of intention. Advances in soft computing algorithms, Research in Computing Science vol. 54, (2011)

6. Clarke, E. M. Jr., Grumberg, O. and Peled. D. A.: Model Checking. MIT Press, Boston, MA., USA, (1999)
7. Cohen, P., Levesque, H.: Intention is choice with commitment. Artificial Intelligence 42(3), 213-261 (1990)
8. Couturat, L.: La logique de Leibniz d'aprés de documents inédits. G. Olms, Hildesheim (1962)
9. 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)
10. Emerson, A.: Temporal and modal logic. In: Handbook of Theoretical Computer Science, Elsevier Science Publishers B.V., Amsterdam, (1990)
11. Dear Bertrand Russell... A Selection of his Correspondence with the General Public 19501968, edited by Barry Feinberg and Ronald Kasrils. London, George Allen and Unwin, (1969)
12. Gabbay, D. M.: Theoretical foundations for nonmonotonic reasoning in expert systems. in K. Apt (ed.), Logics and Models of Concurrent Systems, Berlin and New York: Springer Verlag, pp. 439459 (1985).
13. Governatori, G., Padmanabhan, V. and Sattar, A.: A Defeasible Logic of Policy-based Intentions. In AI 2002: Advances in Artificial Intelligence, LNAI-2557. Springer Verlag (2002)
14. Governatori, G., Terenziani, P.: Temporal Extensions to Defeasible Logic. In Proceedings of IJCAI'07 Workshop on Spatial and Temporal Reasoning, India (2007)
15. A. Guerra-Hernández, J. M. Castro-Manzano, A. El-Fallah-Seghrouchni.: CTLAgentSpeak(L): a Specification Language for Agent Programs. Journal of Algorithms in Cognition, Informatics and Logic, (2009)
16. Hoek, W. van der, Jamroga, W., Wooldridge, M.: Towards a theory of intention revision. Synthese, Springer-Verlag (2007).
17. Konolige, K., Pollack, M. E.: A representationalist theory of intentions. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-93), 390-395, San Mateo: Morgan Kaufmann (1993).
18. Nute, D.: Defeasible logic. In: INAP 2001, LNAI 2543M 151-169, Springer-Verlag, (2003).
19. 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).
20. Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. In D. Gabbay and F. Guenthner (eds.), Handbook of Philosophical Logic, second edition, Vol 4, pp. 219-318. Kluwer Academic Publishers, Dordrecht etc., (2002).
21. 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).
22. 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).
23. Turner, R. and Eden, A.: "The Philosophy of Computer Science", The Stanford Encyclopedia of Philosophy (Summer 2009 Edition), Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/sum2009/entries/computer-science/¿.
24. Wooldridge, M.: Reasoning about Rational Agents. MIT Press, Cambridge (2000).

Appendix

AgentSpeak(L) syntax 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 $AgentSpeak(L)$.

AgentSpeak(L) semantics The operational semantics of $AgentSpeak(L)$ are defined by a transition system, as showed in Figure 1, between configurations $\langle ag, C, M, T, s \rangle$:

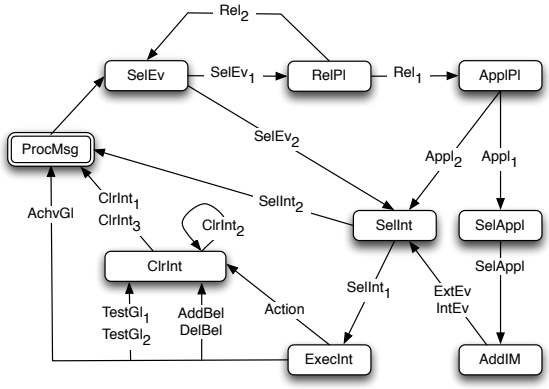


Fig. 1. The interpreter for $AgentSpeak(L)$ as a transition system.

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 the $AgentSpeak(L)$ operational semantics and σ_i, σ_j are agent configurations.