# From Desires to Intentions Through Dialectical Analysis

Nicolas D. Rotstein Alejandro J. Garcia
Guillermo R. Simari
Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET)
Artificial Intelligence Research and Development Laboratory
Department of Computer Science and Engineering
Universidad Nacional del Sur – Bahia Blanca, Argentina
e-mail: {ndr,ajg,grs}@cs.uns.edu.ar

## ABSTRACT

In this work, we introduce a framework where defeasible argumentation is used for reasoning about beliefs, desires and intentions. A dialectical filtering process is introduced in order to obtain a subset of the agent's desires containing only those that are actually achievable in the current situation. In our framework, different agents types can be defined and this will affect the way in which current desires are obtained. Finally, intentions will be current desires that the agent may commit to pursue.

### Categories and Subject Descriptors

I.2.11 [**Artificial intelligence**]: Multiagent systems

### General Terms

Theory

### Keywords

BDI architecture, Defeasible argumentation, Argumentation in agent systems

## 1. INTRODUCTION

In this work, we introduce a framework where defeasible argumentation is used for warranting agent's beliefs, filtering desires, and selecting proper intentions according to a given policy. We allow for the definition of different types of agents, which will affect the way in which desires are filtered.

Autonomous agents based on mental attitudes had gathered special attention in the last years, specially those that follow architectures based on BDI. There are several approaches built upon BDI, some of them introducing new components, like the BOID architecture [2]. Regarding the underlying logic of the sets of *Beliefs*, *Desires* and *Intentions*, some approaches are based on theories such as Default Logic [2, 8], whereas others introduce frameworks combining BDI with an argumentation formalism [1, 5, 6, 7].

In [7] a proposal for using defeasible argumentation to reason about agent's beliefs and desires is introduced. There, they describe a mechanism to filter agent's desires in order to obtain a set of current desires, *i.e.*, those that are achievable in the current situation. However, in this work, agent's intentions are not considered. Here, we extend that approach by adding agent's intentions, and propose a new, more general filtering process that involves the introduction of the notion of *agent type*.

The contribution of our approach is to introduce a BDI architecture that uses a concrete framework based on a working defeasible argumentation system: Defeasible Logic Programming (DeLP). In order to show how an agent can be implemented using this framework, we provide meaningful examples from a robotic soccer domain.

Partially supported by Universidad Nacional del Sur, CONICET, and Agencia Nacional de Promoción Científica y Tecnológica.


*AAMAS'07* May 14–18 2007, Honolulu, Hawai'i, USA.


## 2. WARRANTING BELIEFS

As introduced in [7], agent's beliefs will correspond to the semantics[1] of a *defeasible logic program* $\mathcal{P}_{\mathsf{B}} = (\Pi_{\mathsf{B}},\Delta_{\mathsf{B}})$ [4]. The information that the agent perceives directly from its environment is represented in $\Pi_{\mathsf{B}}$ with a subset of facts denoted $\Phi$. Thus, in the set $\Pi_{\mathsf{B}}$ two disjoint subsets will be distinguished: the subset $\Phi$ of *perceived beliefs* that will be updated dynamically, and a subset $\Pi$ formed with strict rules and facts that will represent static knowledge. Therefore, $\Pi_{\mathsf{B}} = \Phi \cup \Pi$. In addition to the perceived beliefs, the agent may use strict and defeasible rules from $\mathcal{P}_{\mathsf{B}}$ in order to obtain a *warrant* for its *derived beliefs* (see Definition 1).

Since $\Pi_{\mathsf{B}}$ has to be non-contradictory, we assume that perception is correct in the sense that it will not give a pair of contradictory literals. We will also require that no perceived literal in $\Phi$ can be derived directly from $\Pi$. Thus, if $\Pi$ is non-contradictory and these two restrictions are satisfied, then $\Pi_{\mathsf{B}}$ will also be non-contradictory.

**Definition** 1 (BELIEF TYPES). *A* ***perceived belief*** *is a fact in $\Phi$ that represents information that the agent has perceived directly from its environment. A* ***strict belief*** *is a literal that is not a perceived belief, and it is derived from $\Pi_{\mathsf{B}} = \Pi \cup \Phi$ (i.e., no defeasible rules are used for its derivation). A* ***defeasible belief*** *is a warranted literal $L$ supported by an argument $\langle \mathcal{A}, L\rangle$ that uses at least one defeasible rule*

[1]Since the semantics of DeLP is skeptical, there is only one.



978-81-904262-7-5 (RPS) 

*(i.e., $\mathcal{A} \neq \emptyset$). Finally, a* ***derived belief*** *is a strict or a defeasible belief. We will denote with $\mathsf{B}_s$ the set of strict beliefs, and with $\mathsf{B}_d$ the set of defeasible beliefs. Therefore, the beliefs of an agent will be $\mathsf{B} = \Phi \cup \mathsf{B}_s \cup \mathsf{B}_d$.*

**Example** 1. *Consider a robotic-soccer agent Ag that has the following program $(\Pi_\mathsf{B},\Delta_\mathsf{B})$. Here, the set $\Pi_\mathsf{B}$ was divided distinguishing the subset $\Phi$=$\{hasBall(t1), marked(t1)\}$ of perceived facts, and the subset $\Pi$ of non-perceived information:*

$$\Pi = \left\{ \begin{array}{l} mate(t1),\ opponent(o1), \\ (\sim mate(X) \leftarrow opponent(X)), \\ (\sim receive(self) \leftarrow hasBall(self)) \end{array} \right\}$$

$$\Delta_\mathsf{B} = \left\{ \begin{array}{l} (receive(self) \prec hasBall(X), mate(X)), \\ (\sim receive(self) \prec marked(self)), \\ (\sim receive(self) \prec hasBall(X), \sim mate(X)) \end{array} \right\}$$

*In this example, Ag has two perceived beliefs: "player $t1$ has the ball", and "teammate $t1$ is marked". It has also two other facts that are strict beliefs: $mate(t1)$ and $opponent(o1)$. The set $\Pi$ has two strict rules representing that "an opponent is not a teammate" and that "Ag cannot receive the ball from itself"; thus, it infers $\sim mate(o1)$ as a strict belief.*

*The set of defeasible rules $\Delta_\mathsf{B}$ represents that: "if a teammate has the ball, then Ag may receive a pass from it", "being marked is a good reason for not receiving a pass", and "if the one that has the ball is not a teammate, then there are good reasons for not expecting to receive the ball from it". From $(\Pi_\mathsf{B},\Delta_\mathsf{B})$, the argument for $receive(self)$:*
$\{receive(self) \prec hasBall(t1), mate(t1)\}$
*has no defeaters, and therefore, there is a warrant for one defeasible belief: $receive(self)$ (Ag may receive a pass).*

In this approach, we assume a perception function that provides the agent with information about its environment, This function will be invoked by the agent in order to update its perceived beliefs set $\Phi$. When this happens, the new information obtained must override the old one following some criterion. Updating a set of literals is a well-known problem and many solutions exist in the literature [3].

**Example** 2. *Consider the agent of Example 1. If perception is now $\Phi$= $\{hasBall(t1), marked(t1), marked(self)\}$ (i.e., the original situation has changed only in that the agent is now being marked); then, from this new program, the argument for $receive(self)$ has a "blocking defeater", which means that the DeLP answer for both $receive(self)$ and $\sim receive(self)$ will be* UNDECIDED.

*Assume another situation, where $\Phi$= $\{hasBall(o1)\}$. Here, the DeLP answer for $receive(self)$ is* NO, *because there is a warrant for $\sim receive(self)$ supported by the non-defeated argument $\{\sim receive(self) \prec hasBall(o1), \sim mate(o1)\}$.*

## 3. FILTERING DESIRES

Agents desires will be represented by a given set $\mathsf{D}$ of literals, each of which will be a desire that the agent might want to achieve. Note that this set $\mathsf{D}$ may be contradictory. We will assume that beliefs and desires are represented with separate names, *i.e.*, $\mathsf{D} \cap \mathsf{B} = \emptyset$. Hence, a desire cannot be perceived or derived as a belief.

Depending on the situation in which the agent is involved, there could be some desires impossible to be carried out. For example, consider a situation in which the agent does not have the ball and the ball is in a place $p$, then, the desire *shoot* will not be possible to be carried out, whereas $goto(p)$ could be a plausible option. Therefore, agents should reason about their desires in order to select the appropriate ones.

In [7] a reasoning formalism is introduced for selecting from $\mathsf{D}$ those desires that are suitable to be carried out. In order to perform this selection, the agent uses its beliefs (representing the current situation) and a defeasible logic program $(\Pi_F,\Delta_F)$ composed by *filtering rules*. The set of filtering rules represent reasons (for and against) to adopt desires. In other words, filtering rules are devoted to put aside those desires that cannot be achieved in the situation where the agent is involved. A *filtering rule* is a strict or defeasible rule with a desire as head and non-empty body.

**Example** 3. *A robotic-soccer agent could have the set of desires $\mathsf{D} = \{shoot, carry, pass, move\}$ and the following filtering rules:*

$$\Pi_F = \left\{ \begin{array}{l} \sim carry \leftarrow \sim ball \\ \sim shoot \leftarrow \sim ball \\ \sim pass \leftarrow \sim ball \end{array} \right\} \quad \Delta_F = \left\{ \begin{array}{l} shoot \prec goalieAway \\ carry \prec nooneahead \\ pass \prec freeMate \\ \sim shoot \prec farGoal \\ \sim carry \prec shoot \\ move \prec \sim ball \end{array} \right\}$$

Consider a particular situation in which an agent does not have the ball (*i.e.*, $\sim ball \in \Phi$). If the agent has $\Delta_\mathsf{B} = \emptyset$, $\Pi_\mathsf{B} = \Phi$ and the filtering rules $(\Pi_F,\Delta_F)$ from Example 3, then, there are warrants for $\sim carry$, $\sim pass$ and $\sim shoot$. Hence, in this particular situation, the agent should not consider selecting the desires *carry*, *pass*, and *shoot*, because there are justified reasons against them. Observe that these reasons are not defeasible.

Suppose now a new set of perceived beliefs:
$\mathsf{B} = \Phi = \{ball, goalieAway, farGoal\}$,
denoting that the agent has the ball and the opponent goalie is away from its position, but the agent is far from the goal. Then, from the agent's beliefs and the filtering rules $(\Pi_F,\Delta_F)$ of Example 3, there are arguments for both *shoot* and $\sim shoot$. Since these two arguments defeat each other, a blocking situation occurs and the answer for both literals is UNDECIDED. In our approach (as will be explained later) an undecided desire could be eligible.

In this formalism, beliefs and filtering rules are used in combination. Hence, we need to explain how two defeasible logic programs can be properly combined. Agents will have a *de.l.p.* $(\Pi_\mathsf{B},\Delta_\mathsf{B})$ containing rules and facts for deriving beliefs, and a *de.l.p.* $(\Pi_F,\Delta_F)$ with filtering rules for selecting desires. We need to combine these two *de.l.p.*, but the union of them might not be a *de.l.p.*, because the union of the sets of strict rules could be contradictory. To overcome this issue, we use a merge revision operator "$\circ$". Hence, in our case, the join of two *de.l.p.* like $(\Pi_\mathsf{B},\Delta_\mathsf{B})$ and $(\Pi_F,\Delta_F)$ will be a program $(\Pi,\Delta)$, where $\Pi = \Pi_\mathsf{B} \circ \Pi_F$ and $\Delta = \Delta_\mathsf{B} \cup \Delta_F \cup \Delta_X$. We refer the interested reader to [7] for more details and examples of this operator.

**Definition** 2 (CURRENT DESIRES).
*Let $(\Pi_\mathsf{B},\Delta_\mathsf{B})$ be the set containing rules and facts for deriving beliefs; $(\Pi_F,\Delta_F)$, the set carrying filtering rules; and $\Delta_X = \{(\alpha \prec \gamma) \mid (\alpha \leftarrow \gamma) \in (\Pi_\mathsf{B} \cup \Pi_F)$ and $(\Pi_\mathsf{B} \cup \Pi_F) \vdash \{\alpha, \overline{\alpha}\}\}$. Then, let $K = (\Pi_\mathsf{B} \circ \Pi_F,\ \Delta_\mathsf{B} \cup \Delta_F \cup \Delta_X)$ be the knowledge base of an agent. The set $\mathsf{D}^c$ of* Current Desires *is defined as: $\mathsf{D}^c = filter(T, \mathsf{D}, K)$, where the function $filter(\cdot,\cdot,\cdot)$ returns the maximal subset of $\mathsf{D}$ containing those desires that satisfy the selection criterion $T$ from $K$.*

The filtering function can be defined in a modular way. Therefore, different agent types, personalities or behaviours can be obtained depending on the chosen filtering criterion. Agent types using DeLP can be defined as follows:

**Cautious agent:** $filter(T, \mathsf{D}, K) = \{\delta \in \mathsf{D} \mid T(\delta, K) =$ "the answer for $\delta$ from $K$ is YES"$\}$
**Bold agent:** $filter(T, \mathsf{D}, K) = \{\delta \in \mathsf{D} \mid T(\delta, K) =$ "the answer for $\delta$ from $K$ is YES, UNDECIDED or UNKNOWN"$\}$

**Example** 4. *Extending Example 3, if we consider a bold agent as defined above and the set of beliefs:*
$\mathsf{B} = \Phi = \{farGoal, nooneahead, ball\}$, *the agent will generate the set of current desires* $\mathsf{D}^c = \{carry, pass\}$. *Here, we have* $K = (\Phi \circ \Pi_F, \emptyset \cup \Delta_F \cup \emptyset)$. *Regarding the elements in* $\mathsf{D}^c$, *the DeLP answer for shoot is* NO, *for carry is* YES, *and for pass is* UNDECIDED. *Finally, note that a cautious agent would choose carry as the only current desire.*

## 4. SELECTING INTENTIONS

In our approach, an intention will be a current desire that the agent can commit to pursue. To specify under what conditions the intention could be achieved, the agent will be provided with a set of *intention rules*. These rules are denoted $(d \Leftarrow P, C)$, having in its head a literal $d$ that represents a desire that could be selected as an intention, whereas the preconditions part $P$ of its body is a set $\{p_1, \dots, p_n\}$ $(n \geq 0)$ of literals, and the constraints part $C$ is a set $\{c_1, \dots, c_m\}$ $(m \geq 0)$ of literals.

**Example** 5. *A robotic-soccer agent might have the following intention rules:*
$IR_1: \ (carry \Leftarrow \{ball\}, \{\})$, $IR_2: \ (pass \Leftarrow \{ball\}, \{not\ shoot\})$,
$IR_3: \ (shoot \Leftarrow \{ball\}, \{not\ marked\})$,
$IR_4: \ (carry \Leftarrow \{winning\}, \{\})$, $IR_5: \ (move \Leftarrow \{\}, \{\})$

**Definition** 3 (APPLICABLE INTENTION RULE).
*Let* $(\Pi_{\mathsf{B}}, \Delta_{\mathsf{B}})$ *be the set containing rules and facts for deriving beliefs;* $(\Pi_F, \Delta_F)$, *the set carrying filtering rules; and* $\Delta_X = \{(\alpha \prec \gamma) \mid (\alpha \leftarrow \gamma) \in (\Pi_{\mathsf{B}} \cup \Pi_F)$ *and* $(\Pi_{\mathsf{B}} \cup \Pi_F) \vdash \{\alpha, \sim\alpha\}\}$. *Then, let* $K = (\Pi_{\mathsf{B}} \circ \Pi_F,\ \Delta_{\mathsf{B}} \cup \Delta_F \cup \Delta_X)$ *be the knowledge base of an agent, and* $\mathsf{D}^c$, *its set of current desires. Let* $\mathsf{B}$ *be the set of beliefs obtained from* $(\Pi_{\mathsf{B}}, \Delta_{\mathsf{B}})$. *An intention rule* $(d \Leftarrow \{p_1, \dots, p_n\}, \{not\ c_1, \dots, not\ c_m\})$ *is* applicable *iff*
*1.* $d \in \mathsf{D}^c$,
*2. for every precondition* $p_i$ *it holds that* $p_i \in (\mathsf{B} \cup \mathsf{D}^c)$, *and*
*3. for every constraint* $c_j$ *it holds that* $c_j \notin (\mathsf{B} \cup \mathsf{D}^c)$.

**Example** 6. *Consider a bold agent, and* $K$, $\mathsf{B}$ *and* $\mathsf{D}^c$ *as given in Example 4. Now it is possible to determine which of the intention rules of Example 5 are applicable. Rule* $IR_1$ *is applicable because carry* $\in \mathsf{D}^c$. *Rule* $IR_2$ *is applicable because pass* $\in \mathsf{D}^c$, *ball* $\in \mathsf{B}$, *and shoot* $\notin \mathsf{D}^c$.*Rule* $IR_3$ *is not applicable because shoot* $\notin \mathsf{D}^c$. *Rule* $IR_4$ *is not applicable because the precondition is not a literal from* $K$. *Finally,* $IR_5$ *is not applicable because move* $\notin \mathsf{D}^c$. *Thus, the set of applicable rules of the agent is* $\{IR_1, IR_2\}$.

The set of all applicable intention rules contains rules whose heads represent *applicable intentions* that the agent can achieve in the current situation. Depending on the application domain, there are many possibilities for defining a policy to select among a set of applicable intentions.

**Definition** 4 (SET OF SELECTED INTENTIONS).
*Let* $IR$ *be the set of intention rules of an agent, and* $App \subseteq IR$ *be the set of all the applicable intention rules. Let* $p : IR \rightarrow \mathsf{D}$ *be a given function that represents the selection policy. Then, the* set of selected intentions $\mathsf{I}$ *will be* $p(App)$.

In our application domain of robotic soccer, agents have to select a single applicable intention at a time (*i.e.*, an agent cannot shoot and pass the ball at the same time). One possibility for defining a policy that returns a single intention is to provide a sequence with all the intention rules $[IR_1,...,IR_n]$ that represents a preference order among them. Then, the policy $p(App)$ will select the first rule $IR_k$ $(1 \leq k \leq n)$ in the sequence that belongs to *App*, and it will return the head of $IR_k$. In our approach, we consider an agent as a tuple including a set of desires, agent knowledge including perceptions, a set of filtering rules, a filtering function, a set of intention rules, and a policy for selecting intentions.

## 5. CONCLUSIONS

In this paper we have shown how a deliberative agent can represent its perception and beliefs using a defeasible logic program. The information perceived directly from the environment is represented with a subset of perceived beliefs that is dynamically updated, and a subset $\Pi$ formed with strict rules and facts represent other static knowledge of the agent. In addition to this, defeasible argumentation is used in order to warrant agents (derived) beliefs.

Filtering rules have been introduced in order to represent knowledge regarding desires. Defeasible argumentation is used for selecting a proper desire that fits in the particular situation the agent is involved. We allow the representation of different agent types, which will affect the filtering process. In our approach, an intention is a current desire that the agent can commit to pursue. The agent is provided with a set of intention rules that specify under what conditions an intention could be achieved. If there is more than one applicable intention rule, then a policy is used to define a preference criterion among them. Thus, intention policies provide the agent with a mechanism for deciding which intentions should be selected in the current situation.

## 6. REFERENCES


[1] L. Amgoud. A formal framework for handling conflicting desires. In *Proc. of the 7th ECSQARU*, pages 552–563, 2003.
[2] J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre. The boid architecture: conficts between beliefs, obligations, intentions and desires. In *Proc. of 5th Int. Conf. on Autonomous Agents*, pages 9–16. ACM Press, Canada, 2001.
[3] M. Falappa, G. Kern-Isberner, and G. Simari. Belief revision, explanations and defeasible reasoning. *Artificial Intelligence Journal*, 141:1–28, 2002.
[4] A. García and G. Simari. Defeasible logic programming: An argumentative approach. *Theory Practice of Logic Programming*, 4(1):95–138, 2004.
[5] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.
[6] I. Rahwan and L. Amgoud. An argumentation-based approach for practical reasoning. In *Proc. of the 5th AAMAS*, 2006.
[7] N. D. Rotstein and A. J. García. Defeasible reasoning about beliefs and desires. In *Proc. of the 11th NMR*, pages 429–436, 2006.
[8] R. Thomason. Desires and defaults: A framework for planning with inferred goals. In *Proc. of the 7th KR*, pages 702–713, 2000.