Wishful Thinking:
an analysis of conflicts between beliefs and desires

Jan Broersen, Mehdi Dastani and Leendert van der Torre [a]

[a] *Vrije Universiteit*

E-mail: {broersen,mehdi,torre}@cs.vu.nl

In this paper Reiter's default logic is used to study conflicts between conditional beliefs and desires in Thomason's BDP and Broersen *et.al.*'s BOID architecture. In these architectures resolving a conflict is formalized through selection of a subset of the conditionals that does not derive an inconsistency, which is analogous to selecting a subset of rules in input-output logic and selecting an extension in Reiter's default logic. We introduce and discuss several desiderata for this selection process, and we study whether they are satisfied by two specific definitions for belief-desire extensions.

1. Introduction

Thomason [10] proposed an extension of Reiter's default logic with two types of rules, for beliefs and desires. Moreover, he proposed that beliefs override desires to block wishful thinking. In this paper we propose several properties to formalize these intuitions. Thus, we do not introduce a new logic. A crucial desired property is a generalization of the no-wishful thinking principle to extensions. For Thomason's belief-desire case, call the extension of just the beliefs the B extension (defined just like Reiter extensions), and the combined extension of the beliefs and desires the BD extension. The no-wishful thinking principle now says that each BD extension should extend a B extension, and moreover, that each B extension has a BD extension that extends it. We also show that Thomason's logic does not fulfill this desideratum (and neither the existence of extensions desideratum). This property also illustrates how our approach differs from other approaches, like extensions of default logic with preferences and argumentation theory, because in these approaches the conflict resolution mechanism is based on different intuitions and thus on different and incompatible principles, such as for example the weakest link principle (originating from inheritance networks) or the last rule principle (the strength of the argument is the strength of the last rule applied).

The properties investigated in this paper serve as requirements for control loops designed for the BOID architecture [6]. The BOID architecture is an agent architecture consisting of the four components Beliefs, Obligations, Intentions and Desires.

1.1. Reiter's normal default logic

Reiter defined extensions of normal default theories as follows, where we write $\alpha \hookrightarrow w$ for $(\alpha : Mw/w)$ and we write $\langle W, D\rangle$ instead of $\langle D, W\rangle$.

Definition 1. [8, Def. 1] Let $\Delta = \langle W, D\rangle$ be a closed default theory, so that every default of D has the form $\alpha \hookrightarrow w$ where α and w are both closed wffs of a (first-order) language L. For any set of closed wffs $S \subseteq L$ let $T(S)$ be the smallest set of closed formulas from L satisfying the following three properties:

1. $W \subseteq T(S)$
2. $Th_L(T(S)) = T(S)$
3. If $\alpha \hookrightarrow w \in D$, $\alpha \in T(S)$ and $\neg w \notin S$, then $w \in T(S)$.

A set of closed wffs $E \subseteq L$ is an *extension* for Δ iff $T(E) = E$, i.e. iff E is a fixed point of the operator T.

A well-known theorem of Reiter's paper is the following more intuitive characterization of extensions.

Theorem 1. [8, Th. 2.1.] Let $E \subseteq L$ be a set of closed wffs, and let $\Delta = \langle W, D\rangle$ be a closed default theory. Define

$E_0 = W$

and for $i \geq 0$

$E_{i+1} = Th_L(E_i) \cup \{w \mid \alpha \hookrightarrow w \in D \text{ where } \alpha \in E_i \text{ and } \neg w \notin E\}$

Then E is an extension for Δ iff

$E = \cup_{i=0}^{\infty} E_i$.

In this paper we consider Belief-Desire (or BD) theories $\Delta = \langle W, B, D\rangle$, so that every rule of B has the form $\alpha \overset{B}{\hookrightarrow} w$ and every rule of D has the form

$\alpha \overset{D}{\hookrightarrow} w$, where α and w are again closed wffs of L. Note that D no longer stands for Defaults, now it stands for Desires! To facilitate the definitions of the properties in this section we use the following definition.

Definition 2. Let $\Delta = \langle W, B, D\rangle$ be a BD theory. We write $E_R(\Delta)$ for the set of all Reiter extensions of a closed normal default theory and $E_{BD}(\Delta)$ for the set of all extensions of a closed normal BD theory. Moreover, if we consider Reiter extensions of closed normal BD theories consisting of $\alpha \overset{B}{\hookrightarrow} w$ and $\alpha \overset{D}{\hookrightarrow} w$, then we ignore the superscript above the arrows, i.e. we interpret $E_R(\langle W, BD\rangle)$ as $E_R(\langle W, \{\alpha \hookrightarrow w \mid \alpha \overset{B}{\hookrightarrow} w \in BD \text{ or } \alpha \overset{D}{\hookrightarrow} w \in BD\rangle)$. Finally, for representational convenience we write $E_{BD}(W, B, D)$ for $E_{BD}(\langle W, B, D\rangle)$, etc.

In this section we do not define E_{BD} like we defined E_R, but we only consider properties of E_{BD}. In the next section we discuss several definitions of E_{BD} and we consider whether they satisfy the properties or not.

1.2. Classics from Reiter

We first discuss some properties which we have generalized from discussions in Reiter's paper. The first property is called *Existence* and says that there is at least one BD extension, if the facts W are consistent. This is a very desirable and crucial property for agents, because an agent needs an extension to act rationally. Otherwise the agent is stuck or starts to make random movements.

Desired property 1 (Existence). $E_{BD}(W, B, D) \neq \emptyset$ if $\perp \notin Th_L(W)$.

The second property is called *Maximality of extensions* and says that there is no BD extension which is a proper subset of another BD extension.

Desired property 2 (Maximality). For all $E_1, E_2 \in E_{BD}(\Delta)$ we have $E_1 \subseteq E_2$ implies $E_1 = E_2$.

The third property is called *orthagonality of extensions* and says that the union of two distinct BD extensions is inconsistent.

Desired property 3 (Orthagonality). For all $E_1, E_2 \in E_{BD}(\Delta)$ we either have $E_1 = E_2$ or $\perp \in Th_L(E_1 \cup E_2)$.

Intuitions about these first three properties in Reiter's logic carry over to intuitions about these properties in BD theories. Since these properties are well-known from the literature, we do not further discuss them now. More interesting is the following property, called semi-monotonicity, which says that if the sets B and D of $\Delta = \langle W, B, D\rangle$ increase, then its extensions do not shrink or disappear.

Undesired property 1 (Semi-monotonicity). For each $E \in E_{BD}(W, B, D)$, $B' \supseteq B$ and $D' \supseteq D$ there is an $E' \in E_{BD}(W, B', D')$ such that $E' \supseteq E$.

The following example illustrates that semi-monotonicity is incompatible with Thomason's wishful thinking property [10], which states that beliefs override desires. Thomason's example is as follows. If you think it is going to rain and you believe that if it rains, you will get wet, and you wouldn't like to get wet, then you have to conclude that you get wet. Semi-monotonicity is therefore undesirable. We note that semi-monotonicity is also incompatible with non-normal default reasoning and with the introduction of specificity in default logic.

Example 1. Let $\Delta_1 = \langle \emptyset, \emptyset, \{\top \stackrel{D}{\hookrightarrow} p\}\rangle$ and $\Delta_2 = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} \neg p\}, \{\top \stackrel{D}{\hookrightarrow} p\}\rangle$, where $\top$ stands for any tautology. Intuitively we have $E_{BD}(\Delta_1) = \{Th_L(p)\}$, because we apply the only rule available. Moreover, according to Thomason's wishful thinking argument [10] intuitively we have $E_{BD}(\Delta_2) = \{Th_L(\neg p)\}$, because beliefs override desires. This example contradicts the semi-monotonicity property.

We therefore introduce a weaker semi-monotonicity property, which we call *Restricted semi-monotonicity*, that only considers fixed sets of belief rules. The property says that increasing the set of desires D should induce a growth of BD-extensions. This is intuitive, since new desire rules can never overrule existing belief rules. That would be an example of wishful thinking.

Desired property 4 (Restricted semi-monotonicity). For each $E \in E_{BD}(W, B, D)$ and $D' \supseteq D$ there is an $E' \in E_{BD}(W, B, D')$ such that $E' \supseteq E$.

Obviously the restriction of semi-monotoncity to fixed D would be undesired, as illustrated by Example 1, because new beliefs may overrule old desires.

1.3. Wishful thinking

In Example 1 we observed that the semi-monotonicity property implies wishful thinking – desires overriding beliefs – and this property can therefore not be accepted in a beliefs-desire logic. We now look for properties that formalize the opposite, which we call the no-wishful thinking properties, which are properties that guarantee that beliefs override desires. So we want the believes to prevail in conflicts with desires. How can we formulate this intuition as a property of extensions? Desired property 4 does not give us a decision criterion for individual given extensions, since it only says something about what should be the case if they are extended. Therefore we now look at properties that concern the removal of rules. The intuition behind the property of no wishful thinking, that is to be defined next, is the following: If in a conflict between desire and believe the desire rule is removed, the extension cannot increase since the belief rule already had priority over the desire. It seems that no-wishful thinking can be neatly represented by the following *Inverse restricted semi-monotonicity property*, which says that the removal of desires can only decrease the extension, not increase it or remove it.

Undesired property 2 (Inverse restricted semi-monotonicity). For each $E' \in E_{BD}(W,B,D')$ and $D \subseteq D'$ there is an $E \in E_{BD}(W,B,D)$ such that $E \subseteq E'$.

The following example illustrates that this inverse restricted semi-monotonicity property is, unfortunately, too strong.

Example 2. Let $\Delta_1 = \langle \emptyset, \emptyset, \{\top \overset{D}{\hookrightarrow} p\}\rangle$ and $\Delta_2 = \langle \emptyset, \emptyset, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle$. Intuitively we have $E_{BD}(\Delta_1) = \{Th_L(p)\}$ and $E_{BD}(\Delta_2) = \{Th_L(p), Th_L(\neg p)\}$. Thus for $E' = Th_L(\neg p) \in E_{BD}(\Delta_2)$, there is no $E \in E_{BD}(\Delta_1)$ such that $E \subseteq E'$. This example contradicts the inverse restricted semi-monotonicity property.

Example 2 also illustrates where inverse restricted semi-monotonicity goes wrong. The problem of inverse restricted semi-monotonicity is that D may contain rules which have not been used to build E' of $E_{BD}(W,B,D')$, but they may be used when building E of $E_{BD}(W,B,D)$. In the example, this rule was $\top \overset{D}{\hookrightarrow} p$. To eliminate this possibility, we introduce a definition to identify an extension with the set of rules which are applicable in it (sometimes called its generators).

Definition 3 (Applicability). Let $\Delta = \langle W, B, D\rangle$ be a BD theory and let E be one of its extensions. The set of applicable rules $R_B(\Delta, E) = \{\alpha \overset{B}{\hookrightarrow} w \in B | \alpha \wedge w \in E\}$, $R_D(\Delta, E) = \{\alpha \overset{D}{\hookrightarrow} w \in D | \alpha \wedge w \in E\}$, and $R(\Delta, E) = R_B(\Delta, E) \cup R_D(\Delta, E)$.

The following property is called *Generalized no-wishful thinking.* Note that it is a weaker form of the inverse restricted semi-monotonicity property, because $R_D(\langle W, B, D'\rangle, E') \subseteq D'$.

Desired property 5 (Generalized no-wishful thinking). For each $E' \in E_{BD}(W, B, D')$ and $D \subseteq R_D(\langle W, B, D'\rangle, E')$ there is an $E \in E_{BD}(W, B, D)$ such that $E \subseteq E'$.

The following simple examples illustrate no-wishful thinking.

Example 3. Let $\Delta_1 = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \emptyset\rangle$ and $\Delta_2 = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \{\top \overset{D}{\hookrightarrow} q\}\rangle$. If $E_{BD}(\Delta_1) = \{Th_L(p)\}$, then each element of $E_{BD}(\Delta_2)$ has to contain $Th_L(p)$, and $E_{BD}(\Delta_2)$ thus cannot contain for example $Th_L(q \wedge \neg p)$.

Example 4. Let $\Delta_1 = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p\}, \emptyset\rangle$ and $\Delta_2 = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p\}, \{\top \overset{D}{\hookrightarrow} q, p \overset{D}{\hookrightarrow} \neg q\}\rangle$. If $E_{BD}(\Delta_1) = \{Th_L(p)\}$, then each element of $E_{BD}(\Delta_2)$ has to contain $Th_L(p)$, but $E_{BD}(\Delta_2)$ still can contain for example $Th_L(p, q)$ and $Th_L(p, \neg q)$.

The following example illustrates that generalized no-wishful thinking is very strong.

Example 5. Let $\Delta_1 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} \neg q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$ and $\Delta_2 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} \neg q\}, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} q\}\rangle$. Then generalized no-wishful thinking implies that $Th_L(p, q) \notin E_{BD}(\Delta_2)$, because removing $\top \overset{D}{\hookrightarrow} q$ would yield $E_{BD}(\Delta_1) = \{Th_L(p, \neg q)\}$

A simple instance of this generalized no-wishful thinking property, which we call *Restricted no-wishful thinking*, is the case where D is the empty set. This is the definition we already discussed in the introduction, which says that every BD extension extends a B extension.

Desired property 6 (Restricted no-wishful thinking). For each $E' \in E_{BD}(W, B, D')$ there is an $E \in E_{BD}(W, B, \emptyset)$ such that $E \subseteq E'$.

Yet another way to define wishful thinking is to look for a constraint on just the beliefs. The following is a restriction of Maximality (Property 2) to beliefs. Instead of saying that an extension cannot be a strict subset of another extension, we now say that the set of applicable belief rules of one extension cannot be a strict subset of the applicable belief rules of another extension. The intuition is that belief rules should be applied maximally, leaving no room for desires to overrule them.

Undesired property 3 (B maximality). For all $E_1, E_2 \in E_{BD}(\Delta)$ we have $R_B(\Delta, E_1) \subseteq R_B(\Delta, E_2)$ implies $R_B(\Delta, E_1) = R_B(\Delta, E_2)$.

Unfortunately, this definition again does not give intuitive results, as the following example illustrates.

Example 6. Let $\Delta = \langle \emptyset, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle$. Intuitively we have $E_{BD}(\Delta) = \{Th_L(p, q), Th_L(\neg p)\}$. We thus have $R_B(\Delta, Th_L(\neg p)) \subset R_B(\Delta, Th_L(p, q))$. This example contradicts the B maximality property.

The following property is a variant of B maximality.

Undesired property 4 (B). For each $E' \in E_{BD}(W, B, D')$ and $D \subseteq R_D(\langle W, B, D'\rangle, E')$ there is an $E \in E_{BD}(W, B, D)$ such that $R_B(\langle W, B, D\rangle, E) \subseteq R_B(\langle W, B, D'\rangle, E')$.

Example 7 is a generalization of Example 6 that shows why Property 4 is undesired.

Example 7. Let $\Delta_1 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$ and $\Delta_2 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle$. Intuitively we have $E_{BD}(\Delta_1) = \{Th_L(p, q)\}$ and $E_{BD}(\Delta_2) = \{Th_L(p, q), Th_L(\neg p)\}$. We have for $E' = Th_L(\neg p) \in E_{BD}(\Delta_2)$ there is no $E \in E_{BD}(\Delta_1)$ such that $R_B(\Delta_1, E) \subseteq R_B(\Delta_2, E')$.

1.4. Relating BD extension with Reiter extension

We now consider properties that relate the BD extensions to the Reiter extensions. The first property says that a BD extension is a Reiter extension, if one of the sets of rules B or D is empty. Since BD theories are assumed to be extensions of Reiter theories, this property is desired.

Desired property 7 (Reiter). $E_{BD}(W, B, \emptyset) = E_R(W, B)$ and $E_{BD}(W, \emptyset, D) = E_R(W, D)$.

The second property we discuss here is called BD maximality, and says that if you can apply a rule than you do it. That is, you go as far as possible. This property implies that the set of BD extensions are a subset of the set of Reiter extensions where the set of rules consists of the union of belief and desire rules.

Desired property 8 (BD maximality). $E_{BD}(W, B, D) \subseteq E_R(W, B \cup D)$

1.5. Violability

The following example questions the BD maximality property.

Example 8. Let $\Delta = \langle \{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$.[1] We have $E_R(\{\neg q\}, \{p \overset{B}{\hookrightarrow} q, \top \overset{D}{\hookrightarrow} p\}) = \{Th_L(\neg q, p)\}$, and thus with the existence property and the BD maximality property we can derive $E_{BD}(\Delta) = \{Th_L(\neg q, p)\}$. However, $\neg q \wedge p$ implies that to fulfill the desire for p we get into a situation in which something happens which we belief will not happen, namely the exception to the belief that p implies q.

Example 8 is not very convincing, because:

- First, the behavior in Example 8 seems to be what is expected from *conditional* rules. If you do not like it, then you have to formalize the belief rule as $\top \overset{B}{\hookrightarrow} p \rightarrow q$, where $\rightarrow$ is a material implication.
- Second, in the discussion in Example 8 the rules are used as a kind of causal rules. However, if the conditional $p \overset{B}{\hookrightarrow} q$ represents a causal relation, then the world will change such that $\neg q$ will turn into q.

Nevertheless, Example 8 also raises the question whether we can have something less strict than BD maximality. The example suggests that a D rule can only be applied if this does not imply an exception of a B rule. Abnormalities are defined analogous to applicable rules in Definition 3.

[1] If we replace $\neg q$ by $\top \overset{B}{\hookrightarrow} \neg q$ we get an analogous example.

Definition 4 (Abnormality). Let $\Delta = \langle W, B, D\rangle$ be a BD theory and let E be one of its extensions. The set of belief abnormalities is $Ab_B(\Delta, E) = \{\alpha \overset{B}{\hookrightarrow} w \in B | \alpha \wedge \neg w \in E\}$.

The violability property is defined analogous to Desired Property 5, the generalized wishful thinking property. The set of belief abnormalaties (exceptions) may not increase by applying a desire.

Desired property 9 (Violability). For each $E' \in E_{BD}(W, B, D')$ and $D \subseteq R_D(\langle W, B, D'\rangle, E')$ there is an $E \in E_{BD}(W, B, D)$ such that $Ab_B(\langle W, B, D\rangle, E) \supseteq Ab_B(\langle W, B, D'\rangle, E')$.

The following example reconsiders Example 8.

Example 9. Let $\Delta_1 = \langle \{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \emptyset\rangle$ and $\Delta_2 = \langle \{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$. If $Th_L(\neg q, p) \notin E_{BD}(\Delta_1)$ then $Th_L(\neg q, p) \notin E_{BD}(\Delta_2)$.

Example 8 and 9 together illustrate that violability conflicts with BD maximality.

2. BD extensions

We consider two very simple definitions of BD extensions. The first one is called piling, and first tries to apply B rules and thereafter D rules. Once a D rule has been applied, no longer a B rule can be applied.

Definition 5 (Piling). $E_P(W, B, D) = E_R(E_R(W, B), D)$

Proposition 1 (Piling). Properties that hold: existence, maximality, ortagonality, restricted semi-monotonocity, generalized no-wishful thinking, restricted no-wishful thinking, and Reiter. Properties that do not hold: BD maximality, violability.

Proof. The proofs that existence, maximality, ortagonality, and restricted semi-monotonocity hold are all easy generalizations of Reiter's proofs. Generalized (and thus also restricted) no-wishful thinking follows directly from the definitions.

A counterexample to BD maximality:
$\Delta = \langle \emptyset, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$
$E_P(\Delta) = \{Th_L(p)\} \not\subseteq E_R(\Delta) = \{Th_L(p,q)\}$
A counterexample to Violability (Example 9):
$\Delta = \langle \{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$
$E_P(\Delta) = \{Th_L(p, \neg q)\}$

The second simple construction we consider is called *cumulative piling.*

Definition 6 (Cumulative piling). $E_C(W, B, D) = E_R(E_R(B, W), B \cup D)$.

The following proposition shows that we have easily obtained BD maximality, but at a high price: we have lost generalized no-wishful thinking.

Proposition 2. Properties that hold: existence, maximality, ortagonality, restricted semi-monotonocity, restricted no-wishful thinking, Reiter, and BD maximality. Properties that do not hold: generalized no-wishful thinking, violability.

Proof. Most proofs are analogous to the proofs of Proposition 1. BD maximality now obviously holds due to the outer E_R.
Counterexample for generalized no-wishful thinking:
$\Delta_1 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} \neg q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$
$\Delta_2 = \langle \emptyset, \{p \overset{B}{\hookrightarrow} \neg q\}, \{\top \overset{D}{\hookrightarrow} p, p \overset{D}{\hookrightarrow} q\}\rangle$
$E_C(\Delta_1) = \{Th_L(p, \neg q)\}$
$E_C(\Delta_2) = \{Th_L(p, q), Th_L(p, \neg q)\}$
$\{\top \overset{D}{\hookrightarrow} p\} \subseteq R_D(\Delta, Th_L(p,q)) = \{\top \overset{D}{\hookrightarrow} p, p \overset{D}{\hookrightarrow} q\}$

The counterexample to generalized no-wishful thinking illustrates that cumulative piling is not satisfactorily, because after the first round of applying beliefs, it treats beliefs and desires analogously.

3. Concluding remarks

In this paper we have introduced some desired and undesired properties, and checked whether some selection mechanisms satisfied the properties. There are several issue for further research. For conflict resolution we intend to study whether our formal analysis of it can also be used to analyze other approaches. As conflicts are inherent in decision processes, various approaches to handle them have been developed. These approaches vary from specific conflict resolution techniqes to methods to manage the existence of conflicts during a decision process without resolving them (immediately). Examples of approaches of the first kind are [1,9,5,4], where conflicts in application of default rules are resolved by additional (meta-)knowledge on priorities between default rules. Examples of approaches of the second kind can be found in [3,2]. Especially in design processes the existence of conflicts during the whole process is inherent. In [3] it is shown how a component-based approach can increase transparency in conflict management.

There are several technical issues of conflict resolution for beliefs and desires. First, other constructions can be looked for that satisfy for example generalized wishful thinking together with BD maximality, or that satisfy violability. Second, since BD maximality cannot be combined with violability, we could look for an alternative for BD maximality to combine with violability. Third, we will look for further tools to analyze the selection of extensions. BOID ([6]) conflict resolution is based on the three layer structure. First, it is re-interpreted as extension selection, because it means that you take all the effects of your actions into account. Second, some extensions are selected by the prioritization of BOID components, which makes the model simple and opens the way for a list of interesting formal problems and results. Third, the remaining problem is what to do with other conflicts, e.g. between two desires. Here are some suggestions:

- Asking an oracle about propositions, like doing observations in default logic. See [7] for some examples.
- Adding meta-rules like select an extension randomly, first select the beliefs, then select the desires, or choosing desires such that it minimizes the conflicts between beliefs. They all have some drawbacks, but even if the extension selection is problematic, the agent has to choose something.
- Adding some mathematical structure to the rules, with for example priorities as in prioritized default logic, or with penalties and rewards like in decision-theoretic approaches.

References

[1] V.E. Allis, Y.H. Tan, and J. Treur. Meta-level selection techniques for the control of default reasoning. *Future Generation Computer Systems*, 12, Special double issue (2-3): Reflection and Meta-level AI Architectures, (R. Lopez de Mantaras, ed.):189–201, 1996.

[2] F.M.T. Brazier and J. Treur. Managing conflicts in reflective agents. In H.-J. Mueller and R. Dieng, editors, *Computational Conflicts: Conflict Modeling for Distributed Systems*, pages 63–81. Springer Verlag, 2000.

[3] F.M.T. Brazier, P.H.G. Langen van, and J. Treur. Modelling conflict management in design: an explicit approach. *Artificial Intelligence for Engineering Design, Analysis and Manifacturing, (AIEDAM), Special Issue on Conflict Management in Design (I.F.C. Smith, ed.)*, 9, No. 4:353–366, 1995.

[4] G. Brewka. Adding priorities and specificity to default logic. In *Proceedings of the JELIA-94*, 1994.

[5] G. Brewka. Reasoning about priorities in default logic. In *Proceedings of the AAAI-94*, 1994.

[6] J. Broersen, M. Dastani, Z. Huang, J. Hulstijn, and L. van der Torre. The BOID architecture. In *Proceedings of AGENTS'01*, 2001.

[7] J. Lang and P. Marquis. In search of the right extension. In *Proceedings of the KR'2000*. Morgan Kaufmann, 2000.

[8] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 1980.

[9] Y.H. Tan and J. Treur. Constructive default logic and the control of defeasible reasoning. In B. Neumann, editor, *Proc. of the 10th European Conference on Artificial Intelligence, ECAI-92*, pages 299–303. Wiley and Sons, 1992.

[10] R. Thomason. Desires and defaults:. In *Proceedings of the KR'2000*. Morgan Kaufmann, 2000.