Realistic Desires

Jan Broersen* — **Mehdi Dastani**** — **Leendert van der Torre***

* *Vrije Universiteit Amsterdam*
{broersen,torre}@cs.vu.nl
** *Utrecht University*
mehdi@cs.uu.nl
boid *BOID homepage*
http://www.cs.vu.nl/∼boid/

ABSTRACT. *Realism for agents with unconditional beliefs, desires and intentions (BDI agents) has been analyzed in modal logic. This paper provides a logical analysis of realism for agents with* conditional *beliefs and desires in a rule based approach analogous to Reiter's default logic. We distinguish two types of realism, which we call 'a priori' and 'a posteriori' realism. We analyze whether these two new properties are compatible with other properties discussed in the literature, such as existence of extensions. We show that Reiter's default logic is too strong, in the sense that a weaker notion of maximality of extensions is needed to satisfy realism. Finally we show that several existing approaches do not satisfy the new realism properties, and we introduce a new construction that does satisfy them.*

RÉSUMÉ. *A définir par la commande* `\resume{...}`

KEYWORDS: *agent theory, BDI agents, qualitative decision theory, QDT, logic of desires, Reiter's default logic*

MOTS-CLÉS : *A définir par la commande* `\motscles{...}`

1. Introduction

In the BDI (i.e. Belief-Desire-Intention) paradigm [BRA 87, COH 90, RAO 91] the behavior of an agent is governed by the specific way in which it handles the rational balance between its mental attitudes such as beliefs, desires, intentions and obligations. Beliefs are informational attitudes that represent general knowledge about the world as well as knowledge about the agent's environment. Desires and obligations are motivational attitudes that represent wishes and wants, and prohibitions and permissions, respectively. Intentions are attitudes that result from deliberation, representing commitments and previous decisions. The first two attitudes can be related to respectively probabilities and utilities in the classical decision-theoretic approach [LAN 02].

In the BDI approach, the rational balance between mental attitudes is characterized by properties that constrain the interaction between them.

In this paper we are interested in the so-called realism properties. This concerns the question of how an agent's beliefs about the future affect its desires and intentions [WOO 95]. In the unconditional case, this property has been studied by amongst others Cohen and Levecque [COH 90] and Rao and Georgeff [RAO 91]. A distinction has been made between realism ($Bp \rightarrow Dp$, beliefs are desired) and weak realism ($Bp \rightarrow \neg D \neg p$, desires do not conflict with beliefs).[1] In this paper, we say that an agent is realistic if and only if it does not desire states of affairs it believes to be impossible. In other words, our notion of realism is analogous to weak realism. The importance of realism is that a violation of this property may lead to wishful thinking. For example, if the agent believes it is raining but it desires that it is not raining, then the desire should not be used in the agent's decision making process (it should for example not contribute to the derivation of a goal, see below). In the conditional case, realism has been studied by Thomason [THO 00] and Broersen *et al.* [BRO ar]. Whereas the realism property is well understood in the unconditional case, it is much more complex in the conditional one. We illustrate the complications by two examples which play a central role in this paper.

The first example of realism in the context of conditional beliefs and desires illustrates that reasoning with these mental attitudes is related to, but also subtly different from, reasoning with prioritized defaults. The kind of logics discussed in this paper are rule based logics as studied in for example non-monotonic logic, argumentation theory and knowledge based systems. It has been shown that the straightforward local or greedy approach to conflict resolution has counterintuitive consequences. For example, Brewka and Eiter [BRE 99] analyze the following three prioritized default rules (by convention, the lower the number the higher its priority):

$$(1)\ q : \neg p / \neg p \qquad (2)\ \top : p/p \qquad (3)\ \top : q/q$$

The local approach first selects the second rule and thus generates the single preferred extension, $E_1 = Th(\{p, q\})$, generated by the second and third rule, whereas the extension $E_2 = Th(\{\neg p, q\})$ is generated by the first and third rule and is therefore the best choice globally. Now consider this example in a motivational setting in Figure 1.a.

Figure 1.a represents an agent with the following mental attitudes:

1) If q, then the agent believes $\neg p$;

2) The agent desires p;

3) The agent desires q.

1. Realism was introduced by Cohen and Levecque, and Rao and Georgeff introduce weak realism and also a notion of strong realism. For the latter definition they use temporal operators. In particular, if a temporal formula Ep stands for 'there exists a trace in which p holds', then strong realism can be expressed by $DEp \rightarrow BEp$.

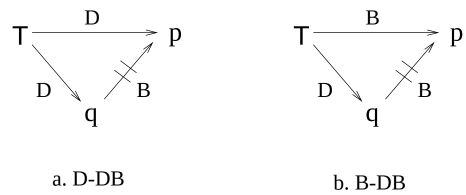


Figure 1. *Conflicts between conditional beliefs and desires*

The realism properties – discussed in detail later in this paper – state that if p is desired, then it is not realistic to desire q. For example, if p is a candidate goal, then q does not contribute to a goal. A realistic agent derives the extension or goal set $E_3 = Th(\{p\})$. Note that E_3 is not an extension in Reiter's sense, because it is not maximal – E_3 is a proper subset of E_1.

The second example illustrates the distinction between two kinds of realism introduced in this paper, which we call 'a priori' realism and 'a posteriori' realism. Both properties are based on the same distinction between the so-called 'a priori' state, in which a certain desire is not taken into account, and the 'a posteriori' state, in which it is taken into account. For example, consider 'a priori' the following set of rules:

1) The agent believes p;

2) If q, then the agent believes $\neg p$.

Moreover, assume that the only thing the agent can deduce from this set and an empty set of observations is p and its logical consequences. In particular, it cannot deduce $\neg q$, because the agent cannot use contraposition. Contraposition is usually forbidden in rule based systems, because otherwise the conditional collapses into material implication (for details consult [MAK 00]). Moreover, consider the following 'a posteriori' rule, leading to the three sentences represented in Figure 1.b:

3) The agent desires q.

The question is now whether the desire for q is realistic. For example, we question whether we may derive a goal for q, i.e. whether striving for q is wishful thinking. The two definitions of realism interpret this example differently. 'A priori' realism says that q is realistic, because in the 'a priori' state we did not believe $\neg q$. 'A posteriori' realism says that q is unrealistic, because in the 'a posteriori' state we have a conflict between two beliefs which we did not have in the 'a priori' state.

The motivation of our work is the formalization of goal generation [BRO ar], although we believe that our notions of realism are also applicable in other contexts. Whereas traditional planning systems take goals as given, in agent systems goals are generated based on motivational attitudes. For example, the agent selects a specific subset of desires, obligations, and intentions as goals. A weakly realistic agent only selects a desire as a goal if the desire does not conflict with the beliefs of the agent. These selected desires are the realistic desires studied in this paper. This can be rephrased in terms of conflict resolution. The goal set of an agent is a non-conflicting subset of the agent's motivational attitudes, and the mechanism through which the conflicts between mental attitudes are resolved characterizes the specific way the agent handles the rational balance between its mental attitudes.

The layout of this paper is as follows. In Section 2 we discuss different kinds of conflicts. In Section 3 we introduce our two kinds of realism, and in Section 4 we analyze the compatibility of these notions with other properties. In Section 5 we check whether several belief-desire logics satisfy realism, including extensions of Reiter's normal default logic such as Thomason's BDP logic in [THO 00] and Broersen *et al.*'s BOID architecture in [BRO ar], and extensions of so-called input/output logics.

2. Realistic desires do not conflict with beliefs

In this section we informally discuss several examples of conflicts in systems with conditional beliefs and desires. Consider the following conflict:

1) The agent believes the car will be sold;
2) The agent believes the car will not be sold.

The agent does not know what to believe: it is confused. Alternatively, the agent has two incompatible belief sets, one which argues that the car will be sold, another which argues that the car will not be sold. Confusion can be formalized by an inconsistent belief set, whereas multiple belief states can be formalized by multiple extensions in for example Reiter's default logic. In this paper we follow the latter approach. Moreover, consider the following conflict:

1) The agent desires the car to be sold;
2) The agent desires the car not to be sold.

The agent has two conflicting desires, which may both become candidate goals. We call this an internal desire conflict. In this paper we again assume that a conflict between desires leads to multiple extensions. Finally, consider the following conflict:

1) The agent believes the car will be sold;
2) The agent desires the car not to be sold.

The agent's desire conflicts with its belief. Such mixed conflicts can be interpreted in various ways. One way, which we adopt in this paper, is due to Thomason [THO 00]. He argues that it is unrealistic to allow the agent's desire to become a goal, and that therefore beliefs should override desires, with the following example. If the agent believes it is raining and it believes that if it rains, it will get wet, and it desires not to get wet, then the agent cannot pursue the goal of not getting wet. This example shows that it is wishful thinking to allow the desire of not getting wet to become a goal. Beliefs prevail in conflicts with desires. Thomason's interpretation can be contrasted with the following example:

1) The agent believes the fence is white;

2) The agent desires the fence to be green.

In this example the agent can see to it that the fence becomes green by painting it, so pursuing the goal that the fence is green is not wishful thinking. The difference between this example and the previous one is that this is not a conflict due to implicit temporal references. The belief implicitly refers to the present whereas the desire refers to the future:

1) The agent believes the fence is white *now*;

2) The agent desires the fence to be green *in the future*.

In this paper we only use abstract examples in which we do not give an interpretation for the propositional atoms. If there is a conflict between a belief and a desire, then there is a real conflict (as in the car selling example), not an apparent conflict (as in the fence example). We also do not discuss the kind of revision or updating involved in the fence example.

The question asked in this paper is how to resolve conflicts between beliefs and desires, in case more than two rules are involved. Consider the example in Figure 2.

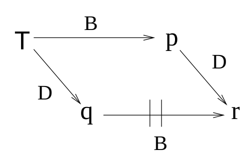


Figure 2. *Is it realistic to desire q? Is it realistic to desire r?*

Figure 2 represents the following four rules:

1) The agent believes p;

2) The agent desires q;

3) If p, then the agent desires r;

4) If q, then the agent believes $\neg r$.

In the following section we give some definitions to determine whether it is realistic to desire q or r.

3. Two notions of realism for conditional desires

In this section we introduce two properties that characterize realistic desires. They are not restricted to one particular logic or architecture, but they can be applied to any extension-based approach.

The reasoning of an agent is characterized by a function, which we denote by Δ, from so-called BD theories (observations with belief and desire rules) to extensions (logically closed sets of propositional sentences that include the observations). This terminology is inspired by Reiter's default logic [REI 80]. However, for now we do not assume any further properties on the relation between BD theories and their extensions. For example, we do not assume that rules are applied to construct extensions.

Definition 1 (BD theory, extension) *A BD theory is a tuple $T = \langle W, B, D \rangle$, where W is a set of propositional sentences of a propositional language L and B and D sets of ordered pairs of such sentences. An extension of T is a logically closed set of L sentences that contains W. Δ is a function which returns for each BD theory a set of its extensions. $\Delta(T)$ is the set of all extensions of a BD theory T (there may be none, one or multiple extensions). We write $Th_L(S)$ for all propositional consequences of the set of propositional formulas S. For representational convenience we write $\Delta(W, B, D)$ for $\Delta(\langle W, B, D \rangle)$, and we write $Th_L(\alpha)$ for $Th_L(\{\alpha\})$.*

Using Definition 1, the example of Figure 2 can be represented by a BD theory $T = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} q, p \overset{D}{\hookrightarrow} r\} \rangle$. Note that Definition 1 allows us to use any pairs of propositional formulas, which means that we consider a more general setting than in the examples thus far.

Realism concerns the rational balance in case of conflict. Therefore we first define what a conflict is. Conflicting theories lead to an inconsistent extension if all applicable rules are applied.[2]

Definition 2 (Conflict) *Let $T = \langle W, B, D \rangle$ be a BD theory. T is a conflict iff there is no consistent logically closed set E of L sentences such that:*

- *$W \subseteq E$*
- *If $a \overset{B}{\hookrightarrow} x \in B$ or $a \overset{D}{\hookrightarrow} x \in D$ and $a \in E$ then $x \in E$*

One way to proceed is to define for each BD theory when a desire is realistic and when it is unrealistic. A drawback of this approach is that it has to commit to a logic of rules for the belief and desire rules. We therefore follow another aprroach, which may be called comparative. The basic pattern is as follows. If a realistic function Δ returns for a BD theory T a set of extensions S, then we can deduce that it does not return for other BD theories T' extensions S'. The latter extensions S' would be unrealistic, i.e. based on unrealistic desires.

2. An alternative stronger definition is that T is a conflict if the following is inconsistent:

$$W \cup \{a \to x \mid a \overset{B}{\hookrightarrow} x \in B \text{ or } a \overset{D}{\hookrightarrow} x \in D\}$$

Which definition of conflict is used depends on the underlying logic of rules, see e.g. [MAK 00] for some possibilities. For the definitions of realism in this paper the exact definition of conflict is not important.

3.1. *A priori realism*

The realism properties are defined in terms of sets of belief and desire rules. However, Δ does not return a set of belief and desire rules, but extensions generated by such rules. We therefore associate with each extension a set of belief and desire rules. The following definition associates an extension with the set of rules which are applied in it (sometimes called its generators [REI 80]).

Definition 3 (Applied rules) *Let $T = \langle W, B, D\rangle$ be a BD theory and let the set E be one of its extensions. The set of applied belief rules in extension E is $R_B(T, E) = \{\alpha \overset{B}{\hookrightarrow} w \in B \mid \alpha \wedge w \in E\}$, and the set of applied desire rules is $R_D(T, E) = \{\alpha \overset{D}{\hookrightarrow} w \in D \mid \alpha \wedge w \in E\}$.*

The intuition behind a priori realism in Property 1 is as follows. Consider a BD theory ($\langle W, B, D\rangle$) and an extension of this BD theory (E) in which at least one desire has been applied. We call this the a posteriori state. We want to ensure that these applied desires are realistic. We therefore consider the state in which this desire has not been applied (BD theory $\langle W, B, D'\rangle$ with extension E'). We call this state the a priori state. We now say that the desire is realistic if the set of applied belief rules in the a priori state is a subset of the set of applied belief rules in the a posteriori state. This implies that the removal of realistic desires from the BD theory cannot lead to the application of belief rules.[3]

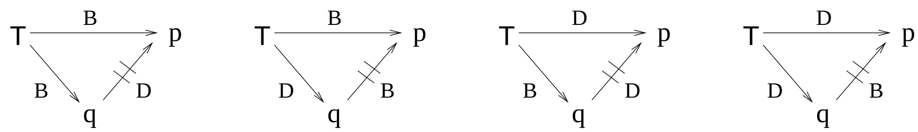
Property 1 (A priori realism) *Δ is a priori realistic iff for each $E \in \Delta(W, B, D)$ and $D' \subseteq R_D(\langle W, B, D\rangle, E)$ there is an $E' \in \Delta(W, B, D')$ such that we have $R_B(\langle W, B, D'\rangle, E') \subseteq R_B(\langle W, B, D\rangle, E)$. We also say that each $E \in \Delta(W, B, D)$ that satisfies the above condition is realistic, and we say that all applied desires of a realistic extension are realistic.*

In the remainder of this section we illustrate a priori realism by some examples. The following triangle example is an extension of the examples discussed in the introduction, because Figure 1.a is Figure 3.d and Figure 1.b is Figure 3.b.

Example 1 *Consider the four triangles in Figure 3. Intuitively, we have a: $\{p, q\}$, b: $\{p\}$ or maybe $\{p, q\}$, c: $\{\neg p, q\}$ or $\{p, q\}$, d: $\{p\}$ or $\{\neg p, q\}$.*

---

3. An alternative closely related definition of a priori realism is as follows. For each $E \in \Delta(W, B, D)$ and $D' \subseteq R_D(\langle W, B, D\rangle, E)$ there is an $E' \in \Delta(W, B, D')$ such that $E' \subseteq E$. This implies that the removal of realistic desires from the BD theory can only decrease the extension, not increase it or remove it. A simple instance of this property, which we may call 'restricted a priori realism,' is the case where D is the empty set. This property says that every BD extension extends a B extension. For each $E \in \Delta(W, B, D)$ there is an $E' \in \Delta(W, B, \emptyset)$ such that $E' \subseteq E$.



a. B-BD b. B-DB c. D-BD d. D-DB

Figure 3. *Desire-belief triangles*

Case a. Let

$$T = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} p, \top \stackrel{B}{\hookrightarrow} q\}, \{q \stackrel{D}{\hookrightarrow} \neg p\}\rangle \text{ and}$$
$$T' = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} p, \top \stackrel{B}{\hookrightarrow} q\}, \emptyset\rangle.$$

T is a conflict, because any set E as defined in Definition 2 contains p as well as $\neg p$. Moreover, assume $\Delta(T') = \{Th_L(p \wedge q)\}$ with $R_B(T', Th_L(p \wedge q)) = \{\top \stackrel{B}{\hookrightarrow} p, \top \stackrel{B}{\hookrightarrow} q\}$. Due to a priori realism we have for each element E of $\Delta(T)$ that $R_B(T, E)$ contains $\{\top \stackrel{B}{\hookrightarrow} p, \top \stackrel{B}{\hookrightarrow} q\}$, and consequently E has to contain $Th_L(p \wedge q)$. $\Delta(T)$ thus cannot contain for example $Th_L(q \wedge \neg p)$. In other words, according to Property 1 we have that the desire for $\neg p$ is unrealistic.

Case b. Let

$$T = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} p, q \stackrel{B}{\hookrightarrow} \neg p\}, \{\top \stackrel{D}{\hookrightarrow} q\}\rangle \text{ and}$$
$$T' = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} p, q \stackrel{B}{\hookrightarrow} \neg p\}, \emptyset\rangle.$$

T is a conflict, because any set E as defined in Definition 2 contains p as well as $\neg p$. Assume $\Delta(T') = \{Th_L(p)\}$ with $R_B(T', Th_L(p)) = \{\top \stackrel{B}{\hookrightarrow} p\}$. Due to a priori realism, each element of $\Delta(T)$ has to contain $Th_L(p)$. Consequently, $\Delta(T)$ can contain $Th_L(p \wedge q)$, but it cannot contain for example $Th_L(q \wedge \neg p)$. In other words, according to Property 1 we have that $\neg p$ is unrealistic. Note that there is not a desire for $\neg p$, but that $\neg p$ would be a believed consequence of a desire (for q). However, it does not imply that the desire for q is unrealistic, an issue discussed again in Example 4 when we have formally introduced a posteriori realism.

Case c. Let

$$T = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} q\}, \{\top \stackrel{D}{\hookrightarrow} p, q \stackrel{D}{\hookrightarrow} \neg p\}\rangle \text{ and}$$
$$T' = \langle \emptyset, \{\top \stackrel{B}{\hookrightarrow} q\}, \emptyset\rangle.$$

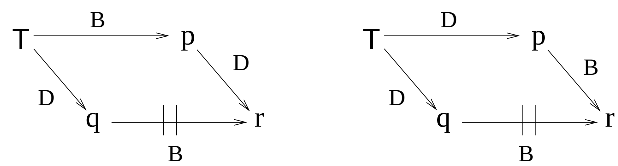
If $\Delta(T') = \{Th_L(q)\}$, then each element of $\Delta(T)$ has to contain $Th_L(q)$, but $\Delta(T)$ still can contain for example $Th_L(p \wedge q)$ and $Th_L(\neg p \wedge q)$. In other words, according to Property 1 neither p nor $\neg p$ would be unrealistic. It illustrates that Property 1 does not classify conflicts between desires as unrealistic.

Case d. Let

$$T = \langle \emptyset, \{q \stackrel{B}{\hookrightarrow} \neg p\}, \{\top \stackrel{D}{\hookrightarrow} p, \top \stackrel{D}{\hookrightarrow} q\}\rangle,$$
$$T' = \langle \emptyset, \{q \stackrel{B}{\hookrightarrow} \neg p\}, \{\top \stackrel{D}{\hookrightarrow} p\}\rangle, \text{ and}$$
$$T'' = \langle \emptyset, \{q \stackrel{B}{\hookrightarrow} \neg p\}, \{\top \stackrel{D}{\hookrightarrow} q\}\rangle.$$

If $\Delta(T') = \{Th_L(p)\}$ and $\Delta(T'') = \{Th_L(q \wedge \neg p)\}$, then a priori realism implies $Th_L(p \wedge q) \notin \Delta(T)$. However, note that $Th_L(p)$ and $Th_L(q \wedge \neg p)$ may be in $\Delta(T)$.

Example 2 considers BD theories with four rules. The first example repeats the example in Figure 2.



a. BD-DB b. DB-DB

Figure 4. *Desire-belief diamonds*

Example 2 *Consider the two diamonds in Figure 4.*

Case a. Let
$$T = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg r\}, \{p \overset{D}{\hookrightarrow} r, \top \overset{D}{\hookrightarrow} q\}\rangle,$$
$$T' = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg r\}, \{p \overset{D}{\hookrightarrow} r\}\rangle, \text{ and}$$
$$T'' = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} q\}\rangle.$$
If $\Delta(T') = \{Th_L(p \wedge r)\}$ and $\Delta(T'') = \{Th_L(p \wedge q \wedge \neg r)\}$ then we have $Th_L(p \wedge q \wedge r) \not\in \Delta(T)$ but $Th_L(p \wedge r)$ and $Th_L(p \wedge q \wedge \neg r)$ may be in $\Delta(T)$ (analogous to Example 1.c).

Case b. Let
$$T = \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} q\}\rangle,$$
$$T' = \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle, \text{ and}$$
$$T'' = \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} q\}\rangle.$$
If $\Delta(T') = \{Th_L(p \wedge r)\}$ and $\Delta(T'') = \{Th_L(q \wedge \neg r)\}$ then we have that the sets $Th_L(p \wedge q \wedge r), Th_L(p \wedge q \wedge \neg r) \not\in \Delta(T)$ but $Th_L(p \wedge r)$ and $Th_L(q \wedge \neg r)$ may be in $\Delta(T)$.

Example 3 considers conflicts between desires. The second BD theory of this example is a variant of the second example discussed in [BRE 99]. Note that this example does not contain conflicts between belief and desire rules, and they thus should not be classified as unrealistic. Example 3 also illustrates that the following two alternative definitions cannot be used to define a priori realism. The first definition considers all desire rules instead of only the applied ones. The second definition maximizes the set of applied belief rules, because they are considered to overrule desire rules.

Alt. 1 For each $E \in \Delta(W, B, D)$ and $D' \subseteq D$ there is an $E' \in \Delta(W, B, D')$ such that $E' \subseteq E$.

Alt. 2 For all $E_1, E_2 \in \Delta(T)$ we have $R_B(T, E_1) \subseteq R_B(T, E_2)$ implies $R_B(T, E_1) = R_B(T, E_2)$.

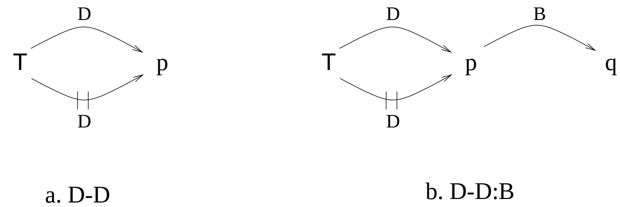


Figure 5. *Conflicting desires*

Example 3 *Consider the conflicts in Figure 5.*

$$\text{Case a. Let } \begin{array}{rcl} T &=& \langle\emptyset,\emptyset,\{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle, \\ T' &=& \langle\emptyset,\emptyset,\{\top \overset{D}{\hookrightarrow} p\}\rangle \text{ and} \\ T'' &=& \langle\emptyset,\emptyset,\{\top \overset{D}{\hookrightarrow} \neg p\}\rangle. \end{array}$$

If we have $\Delta(T') = \{Th_L(p)\}$ and $\Delta(T'') = \{Th_L(\neg p)\}$, then according to Alt. 1 we have for any $E \in \Delta(T)$ that $Th_L(p) \subseteq E$ and $Th_L(\neg p) \subseteq E$. Hence, according to Alt. 1 the only possible extension of T is the inconsistent set. This is obviously counterintuitive. Intuitively we may also have $\Delta(T') = \{Th_L(p)\}$ and $\Delta(T) = \{Th_L(p), Th_L(\neg p)\}$, but this contradicts Alt. 1. However, it does not contradict a priori realism in Property 1, because for $E = Th_L(\neg p)$ we have $R_D(\langle\emptyset,\emptyset,\{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle, E) = \{\top \overset{D}{\hookrightarrow} \neg p\}$, and this set is not a superset of the desire rules in T'. This illustrates that a priori realism is better behaved than Alt. 1.

$$\text{Case b. Let } \begin{array}{rcl} T &=& \langle\emptyset,\{p \overset{B}{\hookrightarrow} q\},\{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} \neg p\}\rangle, \\ T' &=& \langle\emptyset,\{p \overset{B}{\hookrightarrow} q\},\{\top \overset{D}{\hookrightarrow} p\}\rangle \text{ and} \\ T'' &=& \langle\emptyset,\{p \overset{B}{\hookrightarrow} q\},\{\top \overset{D}{\hookrightarrow} \neg p\}\rangle. \end{array}$$

Intuitively we have $\Delta(T) = \{Th_L(p,q), Th_L(\neg p)\}$. This intuition contradicts Alt. 2. due to $R_B(T, Th_L(\neg p)) \subset R_B(T, Th_L(p,q))$ It does not contradict a priori realism in Property 1, because we may have $\Delta(T) = \{Th_L(p,q), Th_L(\neg p)\}$, $\Delta(T') = \{Th_L(p,q)\}$ and $\Delta(T'') = \{Th_L(\neg p)\}$. This illustrates that a priori realism is better behaved than Alt. 2.

3.2. *A posteriori realism*

The second way to define realism is not based on applied rules but on rules which could not be applied, which we call abnormal rules. We only consider abnormal belief rules, not abnormal desire rules.[4]

---

4. An alternative related definition is $Ab_B(T, E) = \{\alpha \overset{B}{\hookrightarrow} w \in B \mid \alpha \in E, w \notin E\}$.

Definition 4 (Abnormal rules) *Let $T = \langle W, B, D \rangle$ be a BD theory and let the set E be one of its extensions. The set of abnormal belief rules is represented by $Ab_B(T, E) = \{\alpha \overset{B}{\hookrightarrow} w \in B \mid \alpha \wedge \neg w \in E\}$.*

A posteriori realism is defined in a similar way as a priori realism in Property 1. It cannot be that a belief rule is abnormal due to an overruling by a desire rule. Thus, removal of desire rules cannot make belief rules normal.

Property 2 (A posteriori realism) *The function Δ is a posteriori realistic if for each $E \in \Delta(W, B, D)$ and $D' \subseteq R_D(\langle W, B, D \rangle, E)$ there is an $E' \in \Delta(W, B, D')$ such that we have $Ab_B(\langle W, B, D' \rangle, E') \supseteq Ab_B(\langle W, B, D \rangle, E)$.*

The following example – already mentioned in the introduction – illustrates that a posteriori realism is different from a priori realism.

Example 4 *Reconsider Example 1.b in Figure 3.b, with*

$$T = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \{\top \overset{D}{\hookrightarrow} q\} \rangle \text{ and}$$
$$T' = \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \emptyset \rangle.$$

Moreover, assume $\Delta(T') = \{Th_L(p)\}$. The extension $E = Th_L(\neg p \wedge q) \in \Delta(T)$ has one abnormal belief rule $\top \overset{B}{\hookrightarrow} p$ which is not an abnormal belief rule for the extension $E' = Th_L(p) \in \Delta(T')$. Moreover, the extension $E = Th_L(p \wedge q) \in \Delta(T)$ has one abnormal belief rule $\top \overset{B}{\hookrightarrow} \neg p$ which is not an abnormal belief rule for the extension E'. Therefore, the rule $\top \overset{D}{\hookrightarrow} q \in D$ is not a realistic desire. Summarizing, if $\Delta(T') = \{Th_L(p)\}$ (no abnormal rules), then $\Delta(T)$ cannot contain for example $Th_L(q \wedge \neg p)$ and moreover, each element of $\Delta(T)$ which contains $Th_L(p)$ cannot contain $Th_L(q)$. In other words, like in Example 1.b we have that $\neg p$ is unrealistic, but in contrast to that example we also have that q is unrealistic.

The analysis of the other examples in Figure 3 remains the same (verification left to the reader).

Example 4 thus illustrates that the two definitions of realism are closely related (for several examples they give the same results), but also that they are subtly different. When choosing one of the two realism properties to guide future development of conflict resolution mechanisms, Examples 1 and 4 may help to choose between Property 1 and 2.[5] The following example further illustrates their difference.

5. On the one hand $p \wedge q$ implies that to fulfi ll the desire for p we get into a state in which something happens which the agent believes that will not happen, namely the exception to the belief that q implies $\neg p$. This argues for Property 2. On the other hand, the behavior in Example 1.b seems to be what is expected from *conditional* beliefs. If you do not like it, then you can replace the second belief rule $q \overset{B}{\hookrightarrow} \neg p$ by the belief rule $\top \overset{B}{\hookrightarrow} (q \rightarrow \neg p)$, where $\rightarrow$ is a material implication. This blocks the inference, but it also derives $\neg q$. This may be an argument for Property 1.

Example 5 *Let* $T = \langle \emptyset, \{a \overset{B}{\hookrightarrow} \neg b\}, \{\top \overset{D}{\hookrightarrow} (a \wedge b)\}\rangle$ *and* $T' = \langle \emptyset, \{a \overset{B}{\hookrightarrow} \neg b\}, \emptyset\rangle$*. If* $\Delta(T') = Th_L(\emptyset)$*, then according to Property 1 we can have that* $Th_L(a \wedge b) \in \Delta(T)$ *but according to Property 2 this is unrealistic.*

The previous two examples illustrated cases in which a certain case was classified as realistic according to Property 1, and as unrealistic according to Property 2. The following example illustrates a case which is classified in the opposite way.

Example 6 *Let* $T = \langle \emptyset, \{a \overset{B}{\hookrightarrow} \neg b\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$ *and* $T' = \langle \emptyset, \{a \overset{B}{\hookrightarrow} \neg b\}, \emptyset\rangle$*. Assume that* $\Delta(T') = \{E'\} = \{\emptyset\}$*. According to Property 1 we can have that* $E = Th_L(a \wedge b \wedge p) \in \Delta(T)$*, because the set of applied belief rules of* E *as well as* E' *are empty. However, according to Property 2 this would be unrealistic, because the set of abnormal belief rules of* E *is* $a \overset{B}{\hookrightarrow} \neg b$ *whereas the set of abnormal belief rules of* E' *is empty.*

In the latter example, proposition $a \wedge b$ cannot be derived from the rules of T. In general, the realism properties we have discussed thus far do not take into account the fact that belief and desire rules are applied to construct the extensions. Whereas there are many ways to construct extensions, there are also some general properties which hold for most rule based systems. They are considered in the following section. In Section 5 we consider existing as well as new conflict resolution mechanisms and test whether they satisfy the realism Properties 1 and 2.

4. Compatibility with other postulates for Δ

In this section we test whether the realism properties are compatible with other properties for rule based formalisms discussed in the literature. In particular we test the compatibility of realism with properties discussed by Reiter [REI 80].

Property 3 says that all extensions returned by Δ are consistent, whenever the observations W are consistent.

Property 3 (Consistency) $Th_L(\bot) \notin \Delta(W, B, D)$ *if* $\bot \notin Th_L(W)$.

Property 4 is called *Existence* and says that Δ returns at least one BD extension. This is a very desirable property for a logic for decision making agents, because an agent needs an extension to act rationally.

Property 4 (Existence) $\Delta(W, B, D) \neq \emptyset$.

Property 5 is called *Rule maximality*. It says that the extensions returned by Δ are such that if a rule can be applied, then it is applied or abnormal. Note that we ignore the superscript above the arrows, if we consider a set that contains belief rules $\alpha \overset{B}{\hookrightarrow} w$ as well as desire rules $\alpha \overset{D}{\hookrightarrow} w$.[6]

Property 5 (Rule maximality) *$\forall E \in \Delta(W, B, D)$ and $\forall \alpha \hookrightarrow w \in B \cup D$, if $\alpha \in E$ then $w \in E$ or $\neg w \in E$.*

Property 6 is called *Constructibility* and says that each extension returned by Δ can be constructed by applying rules as inference rules.

Property 6 (Constructibility) *$\forall E \in \Delta(W, B, D)$ there are $B' \subseteq B$ and $D' \subseteq D$ such that E is the smallest set containing W that is closed under logical consequence and that contains w if it contains α for all rules $\{\alpha \hookrightarrow w \in B' \cup D'\}$.*

The following theorem shows that rule maximality conflicts with realism.

Theorem 1 *There is no Δ that satisfies consistency (Property 3), existence (Property 4), rule maximality (Property 5) and constructibility (Property 6), together with either a priori realism (Property 1) or a posteriori realism (Property 2).*

Proof. *For a priori realism, reconsider Example 2.b.*

$$\begin{array}{lll} \text{Let} \quad T & = & \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} p, \top \overset{D}{\hookrightarrow} q\}\rangle, \\ \quad T' & = & \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle, \text{ and} \\ \quad T'' & = & \langle \emptyset, \{p \overset{B}{\hookrightarrow} r, q \overset{B}{\hookrightarrow} \neg r\}, \{\top \overset{D}{\hookrightarrow} q\}\rangle. \end{array}$$

The only consistent constructible rule maximal extensions of T are $Th_L(p \wedge q \wedge r)$ and $Th_L(p \wedge q \wedge \neg r)$, the only constructible rule maximal extension of T' is $\{Th_L(p \wedge r)\}$, and the only one of $\Delta(T'')$ is $\{Th_L(q \wedge \neg r)\}$. However, as shown in Example 2.b these sets violate the a priori realism property. For T we thus have that all consistent constructible rule maximal extensions are unrealistic.

For a posteriori realism, reconsider Example 1.b and 4.

$$\begin{array}{lll} T & = & \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \{\top \overset{D}{\hookrightarrow} q\}\rangle \text{ and} \\ T' & = & \langle \emptyset, \{\top \overset{B}{\hookrightarrow} p, q \overset{B}{\hookrightarrow} \neg p\}, \emptyset\rangle. \end{array}$$

The only consistent constructible rule maximal extensions of T are $Th_L(q \wedge \neg p)$ and $Th_L(p \wedge q)$, and the only constructible rule maximal extension of T' is $\{Th_L(p)\}$. Neither $Th_L(p \wedge q)$ nor $Th_L(q \wedge \neg p)$ are realistic.

We call a BD logic an extension of Reiter's normal default logic if it selects a subset of the Reiter extensions of the default theory that contains as defaults the union of

6. Note that the minimal (with respect to set inclusion) rule maximal extensions do *not* coincide with Reiter extensions. For example, for a default theory with facts $\emptyset$ and a single default rule $\{\top : p/p\}$, we have that $Th_L(\neg p)$ may be a minimal rule maximal set, but it is not a Reiter extension.

the sets of belief and desire rules. Given that Reiter extensions satisfy consistency, existence, rule maximality, and constructibility, we have the following immediate corollary of Theorem 1.

Corollary 1 *Extensions of Reiter's normal default logic do not satisfy a priori or a posteriori realism.*

The following property is called *Extension maximality* and says that there is no BD extension which is a proper subset of another BD extension.

Property 7 (Extension maximality) *For all $E_1, E_2 \in \Delta(T)$ we have $E_1 \subseteq E_2$ implies $E_1 = E_2$.*

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

Property 8 (Orthogonality) *For all $E_1, E_2 \in \Delta(T)$ we either have $E_1 = E_2$ or $\perp \in Th_L(E_1 \cup E_2)$.*

The following property is called *Semi-monotonicity* and says that if the sets B and D of $\langle W, B, D\rangle$ increase, then its extensions do not shrink or disappear.

Property 9 (Semi-monotonicity) *For each $E \in \Delta(W, B, D)$, $B' \supseteq B$ and $D' \supseteq D$ there is an $E' \in \Delta(W, B', D')$ such that $E' \supseteq E$.*

Most properties (except for rule maximality) we considered are comparative, and are satisfied by the trivial definition $\Delta(W, B, D) = \{Th_L(W)\}$. To discard the trivial definition we ask for overriding in the simplest case.

Property 10 (Basis) $\Delta(\emptyset, \{\top \stackrel{B}{\hookrightarrow} p\}, \emptyset\rangle) = \Delta(\emptyset, \{\top \stackrel{B}{\hookrightarrow} p\}, \{\top \stackrel{D}{\hookrightarrow} \neg p\}\rangle) = \{Th_L(p)\}$ *and* $\Delta(\emptyset, \emptyset, \{\top \stackrel{D}{\hookrightarrow} \neg p\}\rangle) = \{Th_L(\neg p)\}$.

Theorem 2 *Semi-monotonicity (Property 9) conflicts with Basis (Property 10).*

Proof. *We have that* $\Delta(\emptyset, \{\top \stackrel{B}{\hookrightarrow} p\}, \{\top \stackrel{D}{\hookrightarrow} \neg p\}\rangle) = \{Th_L(p)\}$ *and* $Th_L(\neg p) \in \Delta(\emptyset, \emptyset, \{\top \stackrel{D}{\hookrightarrow} \neg p\}\rangle)$ *of Property 10 are a counterexample to Property 9.*

5. Assessment of BD logics

In this section we consider whether several procedures for constructing extensions of BD theories are realistic or not.

5.1. *A simple architecture*

We first consider two definitions of BD extensions based on Reiter's definition of extension (though the first one is not an extension of Reiter's default logic as considered in Corollary 1!). Reiter defines default logics based on first order theories, but we restrict ourselves here to the propositional fragment of his logic. He defines 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 5 *[REI 80, Def. 1] Let $T = \langle W, D\rangle$ be a default theory, so that every default of D has the form $\alpha \hookrightarrow w$ where α and w are both wffs of a (propositional) language L. For any set of wffs $S \subseteq L$ let $F(S)$ be the smallest set of formulas from L satisfying the following three properties:*

1) $W \subseteq F(S)$

2) $Th_L(F(S)) = F(S)$

3) If $\alpha \hookrightarrow w \in D$, $\alpha \in F(S)$ and $\neg w \notin S$, then $w \in F(S)$.

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

We write $\Delta_R(T)$ for the set of all Reiter extensions of a normal default theory. A well-known theorem of Reiter's paper is the following more intuitive characterization of extensions. It is based on a guess of the extension E together with a construction.

Theorem 3 *[REI 80, Th. 2.1.] Let $E \subseteq L$ be a set of wffs, and let $T = \langle W, D\rangle$ be a 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 T iff

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

The first Δ for BD theories we consider is called piling. It first tries to apply belief rules and thereafter desire rules. A belief rule can no longer be applied, once a desire rule has been applied. We write $\Delta_R(\Delta_R(W, B), D)$ for $\bigcup_{E \in \Delta_R(W,B)} \Delta_R(E, D)$.

Definition 6 (Piling) $\Delta_P(W, B, D) = \Delta_R(\Delta_R(W, B), D)$

Theorem 4 (Piling) *Δ_P satisfies a priori realism, but it does not satisfy a posteriori realism. Other properties that hold are consistency, existence, constructibility, extension maximality, orthogonality, and basis, and other properties that do not hold are rule maximality and semi-monotonicity.*

Proof. *The proof that a priori realism holds follows directly from the definitions, and the proof that a posteriori realism does not hold follows from the following example:*

$T = \langle\{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle \quad T' = \langle\{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \emptyset\rangle$
$\Delta_P(T) = \{Th_L(p \wedge \neg q)\} \quad \Delta_P(T') = \{Th_L(\neg q)\}$

The proofs that consistency, existence, constructibility, extension maximality, orthogonality and basis hold are all easy generalizations of Reiter's proofs. A counterexample to rule maximality:

$T = \langle\emptyset, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}\rangle$
$\Delta_P(T) = \{Th_L(p)\}$*, whereas* $p \overset{B}{\hookrightarrow} q$ *can be applied in* $Th_L(p)$

Piling is not satisfactory, because after the first round of applying beliefs, it only considers desires.

5.2. *An architecture with feedback*

The second Δ we consider is called *cumulative piling*.

Definition 7 (Cumulative piling) $\Delta_C(W, B, D) = \Delta_R(\Delta_R(W, B), B \cup D)$.

The following theorem shows that we have obtained rule maximality, but consequently have lost a priori realism.

Theorem 5 (Cumulative piling) Δ_C *does not satisfy any type of realism. Properties that hold are existence, constructibility, extension maximality, orthogonality, rule maximality, and basis, and another property that does not hold is semi-monotonicity.*

Proof. *Most proofs are analogous to the proofs of Theorem 4. Rule maximality now obviously holds due to the outer* Δ_R*, and the absence of realism follows from Theorem 1.*

Cumulative piling is not satisfactory, because after the first round of applying beliefs, it treats beliefs and desires analogously.

5.3. *Broersen* et al.*'s BOID architecture*

The iterative procedure of the Belief-Obligation-Intention-Desire or BOID architecture given in [BRO ar] is presented as an extension of Reiter's more intuitive characterization of extensions in Theorem 3. Like in [MAR 93] it is assumed that there is an order on the rules, represented by ρ. The order on rules reflects some form of overruling of beliefs by desires (and of intentions by beliefs etc.). But it turns out that both a priori and a postoriori realism are not obeyed. It is assumed that the number of rules is finite.[7]

Definition 8 (BOID-realistic agents) *Let $E \subseteq L$ be a set of wffs, and let $T = \langle W, B, O, I, D\rangle$ be a BOID theory with W a set of propositional sentences and B, O, I, and D sets of pairs of such sentences. Moreover, let ρ be a realistic function from the rules of $B \cup O \cup I \cup D$ to the integers iff it associates with each rule a unique number, such that $\rho(r_1) < \rho(r_2)$ if $r_1 \in B$ and $r_2 \in O \cup I \cup D$.*

Given a function ρ, define

$E_0 = W$

and for $i \geq 0$

$\rho_{i+1} = \min\{\rho(\alpha \hookrightarrow w) \mid \alpha \hookrightarrow w \in B \cup O \cup I \cup D, \alpha \in E_i$ *and* $w, \neg w \notin E_i\}$

$E_{i+1} = Th_L(E_i \cup \{w \mid \alpha \hookrightarrow w \in B \cup O \cup I \cup D$ *where* $\rho(\alpha \hookrightarrow w) = \rho_{i+1}\})$ *if a minimal element exists,* $E_{i+1} = E_i$ *otherwise*

Then E is an extension for T iff there exists a realistic function ρ such that

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

We write $\Delta_{BOID}(T)$ for all extensions of $T = \langle W, B, \emptyset, \emptyset, D\rangle$.

Theorem 6 (BOID) *Δ_{BOID} does not satisfy any type of realism. Properties that hold are consistency, existence, rule maximality, contructibility, extension maximality, orthogonality, and basis, and another property that does not hold is semi-monotonicity.*

Proof. *Proofs are analogous to the proofs of Theorem 4. Rule maximality now holds due to iteration, and the absence of realism follows again from Theorem 1.*

Given the negative results on a greedy approach in non-monotonic reasoning (as mentioned in the introduction, see [BRE 99]), it does not come as a surprise that the BOID architecture as proposed in [BRO ar] does not satisfy realism as defined in this paper. It is a consequence of the fact that this BOID architecture satisfies rule maximality. However, note that rule maximality only holds in the limit, i.e. when full extensions are generated. In practice only partial extensions may be generated in the BOID architecture (due to limited resources) and a greedy approach may still be preferred.

7. For infinite sets of rules things get more complicated, because after ω steps we may only have applied the first round of beliefs.

5.4. *Thomason's BDP logic*

Thomason gives two definitions of his extensions, a fixed point definition like in Definition 5 and an iterative one like the construction in Theorem 3. He also claims that these two definitions are equivalent, just like Definition 5 is equivalent to the construction in Theorem 3. However, they differ on at least two parts, namely on the cases in which defaults are overridden as well as the notion of extension. The first definition [THO 00, Def. 2.4-2.5] does not satisfy the Basis property (Property 10),[8] and thus has unintended consequences. We therefore consider Thomason's second definition. Again we only consider the propositional fragment.

Definition 9 (BDP) *[THO 00, Def. 2.1 - 2.3, 2.6 - 2.7] Let $T = \langle W, B, D\rangle$, and $\vdash$ is the consequence relation of propositional logic, $Th_L(F) = \{A \mid F \vdash A\}$. Applicability is defined relative to two parameters: a set F of premises and a "conjectured extension" F^* that is used to test consistency in applying rules.*

1) Applicability for belief rules. A belief rule $\alpha \overset{B}{\hookrightarrow} w$ is applicable to F relative to F^*, where F and F^* are sets of formulas, iff $F \vdash \alpha$ and $F^* \not\vdash \neg w$. $\alpha \overset{B}{\hookrightarrow} w$ is vacuously *applicable to F relative to F^* if it is applicable to F relative to F^* and $w \in T$.*

2) B-conflictedness for desire rules. $\alpha \overset{D}{\hookrightarrow} w$ is B-conflicted for F with respect to F^*, T iff for some $\alpha_1 \overset{B}{\hookrightarrow} w_1, \ldots, \alpha_n \overset{B}{\hookrightarrow} w_n \in B$, $F \vdash \alpha_i$ for all i, $1 \leq i \leq n$ and $F \cup \{w_1, \ldots w_n\} \vdash \neg w$.

3) Applicability for D-rules. A desire rule $\alpha \overset{D}{\hookrightarrow} w$ of $\langle W, B, D\rangle$ is applicable to F, relative to F^* and T, if (1) $F \vdash \alpha$ and $F^* \not\vdash \neg w$, and (2) $\alpha \overset{D}{\hookrightarrow} w$ is not B-conflicted for F with respect to F^*. $\alpha \overset{D}{\hookrightarrow} w$ is vacuously applicable *to F relative to F^* if it is applicable to F relative to F^* and $w \in F$.*

$P(T, E)$ is the sequence $\{E_i^{P(T,E)} \mid i \in \omega\}$ defined as follows:

1) $E_0^{P(T,E)} = W$

2) $E_{i+1}^{P(T,E)} = Th_L(E_i^{P(T,E)} \cup \{w\})$ if there is a rule in $B \cup D$ that is nonvacuously applicable to $E_i^{P(T,E)}$ relative to T, E, where the alphabetically first such rule has the form $\alpha \overset{B}{\hookrightarrow} w$ or $\alpha \overset{D}{\hookrightarrow} w$.

3) $E_{i+1}^{P(T,E)} = E_i^{P(T,E)}$ if no rule in B or D is nonvacuously applicable to $E_i^{P(T,E)}$ relative to T, E.

E is a BD extension of T if $E = lim(P(T, E)) = \cup\{E_i^{P(T,E)} \mid i \in \omega\}$. We write Δ_{BDP} for the set of all BD extensions of T.

---

8. The following example illustrates that the extensions based on BDP fixpoints (which we denote with BDP-fp) do not satisfy the Basis property:

$\Delta_{BDP-fp}(\emptyset, \{\top \overset{B}{\hookrightarrow} p\}, \emptyset) = \{Th_L(p), Th_L(\neg p)\} \not\subseteq \{Th_L(p)\}$.

The main distinction between Δ_{BOID} and Δ_{BDP} is that in the BOID conflict resolution is defined in terms of the order in which rules are applied, while in BDP desire rules are tested for conflicts with belief rules.

Theorem 7 (BDP) *Δ_{BDP} does not satisfy any type of realism. Properties that hold are consistency, existence, constructibility, extension maximality, ortogonality, and basis, and other properties that do not hold are rule maximality and semi-monotonicity.*

Proof. *Most proofs are analogous to the proofs of Theorem 4. The following example illustrates that a priori realism does not hold:*

$\Delta_{BDP}(\emptyset, \{\top \overset{B}{\hookrightarrow} p, p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} \neg q\}) = \{Th_L(p \wedge q), Th_L(p \wedge \neg q)\}$

$\Delta_{BDP}(\emptyset, \{\top \overset{B}{\hookrightarrow} p, p \overset{B}{\hookrightarrow} q\}, \emptyset) = \{Th_L(p \wedge q)\}$

The following example illustrates that a posteriori realism does not hold:

$\Delta_{BDP}(\{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \{\top \overset{D}{\hookrightarrow} p\}) = \{Th_L(p \wedge \neg q)\}$

$\Delta_{BDP}(\{\neg q\}, \{p \overset{B}{\hookrightarrow} q\}, \emptyset) = \{Th_L(\neg q)\}$

The following example illustrates that rule maximality does not hold, because all desires are overruled in case of a conflict between two beliefs:

$\Delta_{BDP}(\emptyset, \{\top \overset{B}{\hookrightarrow} p, \top \overset{B}{\hookrightarrow} \neg p\}, \{\top \overset{D}{\hookrightarrow} q\}) = \{Th_L(p), Th_L(\neg p)\}$

Theorem 1 suggests that rule maximality is the main cause for lack of realism. However, despite the fact that rule maximality does not hold, Δ_{BDP} does not satisfy realism. Thomason introduces on his website an updated version of the theory introduced in [THO 00]. However, the present version (May 2002) also does not satisfy the realism properties.

5.5. *Revision of rule sets*

The final Δ we discuss does not select a subset of the Reiter extensions, like Δ_C and Δ_{BOID}, but it may select more extensions. In particular, it may select extensions for which rule maximality does not hold (cf. [MAK 01]).

Definition 10 (BDA) *Let L be a propositional language, let $T = \langle W, B, D\rangle$ be a BD theory with W a subset of L and B and D sets of ordered pairs of L written as $\alpha \hookrightarrow w$. Moreover, for a rule set $R \subseteq B \cup D$ we say that:*

– $E_a(W, R)$ is the unconstrained application of rules of R to W, i.e. it is the smallest set E satisfying the conditions $E \subseteq L$, $Th_L(E) = E$, $W \subseteq E$, and if $a \hookrightarrow x \in R$ and $a \in E$, then $x \in E$.

– $CF(W, R)$ holds, i.e. R is conflict free given W, iff $E_a(W, R)$ is consistent (analogous to Definition 2).

– *$MCF(W, R)$ is the set of maximal conflict free subsets of R given W if such sets exist, $\{\emptyset\}$ otherwise, i.e.:*

if W is consistent
then $MCF(W, R) = \{R' \subseteq R \mid CF(W, R')$ and $\nexists R'' : R' \subset R''$ and $R'' \subseteq R$ and $CF(W, R'')\}$
else $MCF(W, R) = \{\emptyset\}$.

$\Delta_a(W, R)$ is the corresponding set of extensions:

$$\Delta_a(W, R) = \{E_a(W, R') \mid R' \in MCF(W, R)\}$$

$\Delta_A(W, B, D)$ is $\Delta_a(\Delta_a(W, B), D)$ for $\bigcup_{E \in \Delta_a(W,B)} \Delta_a(E, D)$.

$$\Delta_A(W, B, D) = \Delta_a(\Delta_a(W, B), D)$$

Due to the construction we have in this context no distinction between piling and cumulative piling, i.e.:

$$\Delta_A(W, B, D) = \Delta_a(\Delta_a(W, B), D) = \Delta_a(\Delta_a(W, B), B \cup D)$$

Moreover, we have the following result.

Theorem 8 *Δ_A satisfies a priori and a posteriori realism in Property 1 and 2. Other properties that hold are consistency, existence, constructibility, and basis, and other properties that do not hold are rule maximality, extension maximality, ortogonality, and semi-monotonicity.*

Proof. *Consider a priori realism in Property 1. Let E be an extension of $T = \langle W, B, D\rangle$. Moreover, assume a set of desire rules $D' \subseteq R_D(T, E)$. We have to show that there is an extension E' of $T' = \langle W, B, D'\rangle$ such that $R_B(T', E') \subseteq R_B(T, E)$. We prove the existence of such an extension by constructing one.*

– If W is inconsistent, then E is inconsistent and the property holds. So assume W is consistent.

– Since $E \in \Delta_A(T)$, there must exist $B^+ \in MCF(W, B)$ and $D^+ \in MCF(E_a(W, B^+), D)$ such that $E = E_a(W, B^+ \cup D^+)$.

– Let $E' = E_a(W, B^+ \cup D')$. We have $E' \subseteq E$ due to monotonicity of E_a in R. Moreover, $E' \in \Delta(T')$, because (1) $B^+ \in MCF(W, B)$ due to construction of E and (2) $D' \in MCF(E_a(W, B^+), D')$, because D' is maximal and E' is consistent due to $E' \subseteq E$ and consistency of E.

– We have $R_B(T', E') \subseteq R_B(T, E)$ due to monotonocity of R_B in E.

For a posteriori realism, we can repeat the same construction and prove in the final step:

– We have $Ab_B(T', E') \supseteq Ab_B(T, E)$ due to monotonocity of Ab_B in E.

We have the following corollaries of Theorem 8.

Corollary 2 *The realism properties are not internally inconsistent.*

Corollary 3 *The realism properties are not mutually exclusive.*

Corollary 4 *Extensions of input/output logics [MAK 00, MAK 01] can satisfy a priori and a posteriori realism.*

6. Further research

The formal analysis of conflicts between attitudes leaves several issues for further research. First, realism properties can be generalized to the multiple attitudes case. We conjecture that the overriding of desires by beliefs can be generalized to more complex situations, such as the following example:

1) if a then the agent believes b

2) the agent intends $\neg b$

3) the agent desires a

Using the notion of a BOID theory (see Definition 8), its set of extensions can be determined as $\Delta_{BOID}(\emptyset, \{a \overset{B}{\hookrightarrow} b\}, \emptyset, \{\overset{I}{\hookrightarrow} \neg b\}, \{\top \overset{D}{\hookrightarrow} a\}\rangle$.

Second, the realism properties can be generalized to other input/output logics [MAK 00, MAK 01] than Reiter's normal default logic. Although the properties are defined independent of the logic, both Definition 3 and 4 of applied and abnormal rules must be adapted if we allow other logics. For example, with reasoning by cases we may have $\Delta(\emptyset, \{\alpha \overset{B}{\hookrightarrow} w, \neg\alpha \overset{B}{\hookrightarrow} w\}, \emptyset) = \{Th_L(w)\}$. Moreover, we may allow that the observations are not included in the extensions.

Third, other constructions can be looked for that satisfy a priori or a posteriori realism.

Fourth, other postulates can be studied, for example to distinguish Δ_C and Δ_{BOID}. Moreover, since rule maximality cannot be combined with realism, we could look for an alternative for rule maximality to combine with realism.

7. Summary

In this paper we introduce two notions of realism, called a priori and a posteriori realism. We show that the properties are consistent and can be combined (Corollary 2 and 3). We study the compatibility of realism with other properties discussed in the literature, and we show that extensions of Reiter's default logic (in the sense that they select a subset of the Reiter extensions) cannot satisfy the realism properties

(Corollary 1) in contrast to extensions of input/output logics (Corollary 4). We also test the properties in several existing as well as new conflict resolution mechanisms and we show that there exist mechanisms that satisfy them.

Acknowledgment

Thanks to Zisheng Huang and Joris Hulstijn for discussions on the issues discussed in this paper.

8. References

[BRA 87] BRATMAN M., *Intention, plans, and practical reason*, Harvard University Press, Cambridge Mass, 1987.

[BRE 99] BREWKA G., EITER T., "Preferred answer sets for extended logic programs", *Artificial Intelligence*, vol. 109 (1-2), 1999, p. 297-356.

[BRO ar] BROERSEN J., DASTANI M., HULSTIJN J., VAN DER TORRE L., "Goal generation in The BOID architecture", *Cognitive Science Quarterly*, , To appear.

[COH 90] COHEN P., LEVESQUE H., "Intention is choice with commitment", *Artificial Intelligence*, vol. 42, 1990, p. 213–261.

[LAN 02] LANG J., VAN DER TORRE L., WEYDERT E., "Utilitarian Desires", *Autonomous Agents and Multi Agent systems*, vol. 5:3, 2002, p. 329-363.

[MAK 00] MAKINSON D., VAN DER TORRE L., "Input-output logics", *Journal of Philosophical Logic*, vol. 29, 2000, p. 383-408.

[MAK 01] MAKINSON D., VAN DER TORRE L., "Constraints for input-output logics", *Journal of Philosophical Logic*, vol. 30, 2001, p. 155-185.

[MAR 93] MAREK V., TRUSZCZYNSKI M., *Nonmonotonic logic: Context-dependent reasoning*, Springer, Berlin, 1993.

[RAO 91] RAO A., GEORGEFF M., "Modeling Rational Agents within a BDI-Architecture", ALLEN J., FIKES R., SANDEWALL E., Eds., *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, 1991, p. 473–484.

[REI 80] REITER R., "A logic for default reasoning", *Artificial Intelligence*, vol. 13, 1980, p. 81–132.

[THO 00] THOMASON R. H., "Desires and Defaults: A Framework for Planning with Inferred Goals", COHN A., GIUNCHIGLIA F., SELMAN B., Eds., *KR2000: Principles of Knowledge Representation and Reasoning*, San Francisco, 2000, Morgan Kaufmann, p. 702–713.

[WOO 95] WOOLDRIDGE M., JENNINGS N., "Intelligent agents: theory and practice", *The Knowledge Engineering Review*, vol. 10:2, 1995, p. 115–152.