Rationality Postulates: applying argumentation theory for non-monotonic reasoning

Martin Caminada
Cardiff University
CaminadaM@cardiff.ac.uk

Abstract

The current review paper examines how to apply Dung's theory of abstract argumentation to define meaningful forms of non-monotonic inference. The idea is that arguments are constructed using strict and defeasible inference rules, and that it is then examined how these arguments attack (or defeat) each other. The thus defined argumentation framework provides the basis for applying Dung-style semantics, yielding a number of extensions of arguments. As each of the constructed arguments has a conclusion, an extension of arguments has an associated extension of conclusions. It are these extensions of conclusions that we are interested in. In particular, we ask ourselves whether each of these extensions is (1) consistent, (2) closed under the strict inference rules and (3) free from undesired interference. We examine the current generation of techniques to satisfy these properties, and identify some research issues that are yet to be dealt with.

1 Introduction

Argumentation, as it takes place in everyday life, is never completely abstract. Commonly, arguments are exchanged in order to determine what to do or what to believe. These arguments tend to be composed of reasons, some of which are strict and some of which are defeasible. Strict reasons (like rules of logic) provide conclusive evidence for a claim (like "Socrates is a man. All men are mortal. Therefore, Socrates is mortal.") whereas defeasible reasons (like rules of thumb) provide evidence for their claim that is only valid in the absence of counter evidence (like "Tux is a bird. Therefore Tux can fly."). The existence of defeasible reasons illustrates that for

The current paper will also be published as a chapter in the Handbook of Formal Argumentation.

commonsense reasoning, classical logic is often not sufficient, and that some form of nonmonotonic reasoning (as for instance provided by formal argumentation theory) is necessary.

Whereas defeasible reasons (formally represented as defeasible rules) provide a basis for nonmonotonic reasoning, strict reasons (formally represented as strict rules) provide the ability to model hard constraints (like "given our budget, if we acquire both product X and Y, then we cannot acquire product Z anymore"). By doing so, strict rules provide an important aspect of commonsense reasoning: the ability to reason about an outside world that has particular constraints (for instance of physical or financial nature) that are not subject to discussion.[1]

Suppose one would like to apply Dung's theory in the presence of strict and defeasible rules. That is, the idea is to apply the strict and defeasible rules to construct the arguments of the argumentation framework.[2] How can one be sure that the outcome makes sense from a logical perspective? Suppose there exists a rule representing the reason "given the current budget, if we acquire both product X and Y, then we cannot acquire product Z anymore", together with various other rules. In that case, what one would like to avoid is arguments for buying product X, Y and Z becoming justified (perhaps even in the same extension) because this would mean the constraint is violated. In principle, we could of course look inside of the arguments to check that what we select does not violate any constraint. However, the whole idea of Dung's abstract argumentation theory[3] is *not* to look at the internal structure of the arguments, and to select them based purely on their position in the graph. However, if one cannot look inside of the arguments when selecting them, then how does one make sure that the overall outcome (regarding conclusions on, say, what to do or what to believe) makes any sense?

In the current paper, we examine the question of how to apply Dung's theory

[1]Some argumentation researchers have claimed (personal communication) that if one digs deep enough, even strict rules start to have exceptions, and that therefore only defeasible rules exist. While this may be true from a philosophical perspective, one often wants to restrict the domain of reasoning and not take the more esoteric exceptions into account. The rule "given the current budget, if we acquire both product X and Y, we cannot acquire product Z anymore" may have exceptions if one is willing to steal, but this exception is of little relevance when the setting is a meeting at work. Also, the very idea of modelling information (be it by means of rules or by any other means) is that one limits oneself to a particular Universe of Discourse. Hence, strict rules can be seen as defeasible rules whose exceptions are beyond our current Universe of Discourse.

[2]Basically, this is done by chaining the rules together into inference trees, like is for instance done in [Modgil and Prakken, 2014; Toni, 2014; Caminada *et al.*, 2014b; Caminada *et al.*, 2015].

[3]Keep in mind that in Dung's theory, arguments are abstract, not atomic. Atomic would mean that arguments have no internal structure at all. Abstract means that arguments do have an internal structure, but that one does not take this structure into account (that is, one has *abstracted* from the internal structure).

of abstract argumentation for the purpose of non-monotonic reasoning with strict and defeasible rules. That is, we examine how to apply abstract argumentation semantics while making sure the overall outcome (in terms of justified conclusions) still makes sense. The remaining part of this paper is structured as follows. First, we will state some formal preliminaries on rule-based argumentation in Section 2. Then, in Section 3 we examine three desirable properties of the overall outcome (direct consistency, indirect consistency and closure) and examine various ways of satisfying these properties. Then, in Section 4 we examine two additional desirable properties (non-interference and crash resistance) that are particularly relevant when the strict rules are derived from classical logic, and again examine various ways of satisfying these properties. We round off with a summary and discussion in Section 5.

2 Formal Preliminaries

In the current section, we outline the process of constructing an argumentation framework from a set of strict and defeasible rules. For current purposes, we base our approach on the work of Caminada *et al.* [2014b].[4]

Definition 1. *Given a logical language that is closed under negation ($\neg$), an* argumentation system *is a tuple $AS = (\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ where:*

- *$\mathcal{R}_s$ is a finite set of strict inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ (where φ_i, φ are meta-variables ranging over $\mathcal{L}$ and $n \geq 0$)*
- *$\mathcal{R}_d$ is a finite set of defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ (where φ_i, φ are meta-variables ranging over $\mathcal{L}$ and $n \geq 0$)*
- *$\mathtt{n}$ is a partial function such that $\mathtt{n} : \mathcal{R}_d \longrightarrow \mathcal{L}$*
- *$\leq$ is a partial pre-order on $\mathcal{R}_d$*

We write $\psi = -\varphi$ in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$ (we will sometimes informally say that formulas φ and $-\varphi$ are each other's negation).

To keep things simple, we assume that the logical language $\mathcal{L}$ consists of literals only.[5]

[4] As such, we will for instance not consider the notion of contraries [Modgil and Prakken, 2014] or any other notions in ASPIC+ that are not relevant for current purposes.

[5] In Section 4 we generalise things by having $\mathcal{L}$ be the language of propositional logic.

In the following definition, arguments are constructed from strict and defeasible rules in an inductive way. This process starts from the strict and defeasible rules with empty antecedents (so where $n = 0$).

Definition 2. *An* argument *A on the basis of an argumentation system $AS = (\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ is defined as:*

1. *$A_1, \ldots, A_n \rightarrow \psi$ if $A_1 \ldots A_n$ ($n \geq 0$) are arguments, and there is a strict rule $Conc(A_1), \ldots, Conc(A_n) \rightarrow \psi$ in $\mathcal{R}_s$. In that case we define*
   $Conc(A) = \psi$,
   $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$.
   $DefRules(A) = DefRules(A_1) \cup \ldots \cup DefRules(A_n)$,
   $TopRule(A) = Conc(A_1), \ldots, Conc(A_n) \rightarrow \psi$

2. *$A_1, \ldots, A_n \Rightarrow \psi$ if $A_1 \ldots A_n$ ($n \geq 0$) are arguments, and there is a defeasible rule $Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$. In that case we define*
   $Conc(A) = \psi$,
   $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$,
   $DefRules(A) = DefRules(A_1) \cup \ldots \cup DefRules(A_n) \cup$
   $\{Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi\}$,
   $TopRule(A) = Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi$.

Furthermore, for any argument A and set of arguments E:

- *A is* strict *iff $DefRules(A) = \emptyset$;* defeasible *iff $DefRules(A) \neq \emptyset$;*
- *If $DefRules(A) = \emptyset$, then $LastDefRules(A) = \emptyset$, else;*
  if $A = A_1, \ldots, A_n \Rightarrow \phi$ then $LastDefRules(A) = \{Conc(A_1), \ldots, Conc(A_n) \Rightarrow \phi\}$, otherwise $LastDefRules(A) = LastDefRules(A_1) \cup \ldots \cup LastDefRules(A_n)$.
- *$Concs(E) = \{Conc(A) \mid A \in E\}$*
- *The closure under strict rules of E, denoted $Cl_S(E)$ is the smallest set containing $Concs(E)$ and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedent is contained in $Cl_S(E)$.*

For current purposes (as well as is done in [Caminada and Amgoud, 2007; Prakken, 2010; Caminada *et al.*, 2014b]) we assume that the set of strict rules is consistent in the following way.

Definition 3. *Let $AS = (\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ be an argumentation system. We say that AS and $\mathcal{R}_s$ are consistent iff no strict arguments A and B exist such that $\mathtt{Conc}(A) = -\mathtt{Conc}(B)$*

Definition 4. *Let A and B be arguments. We say that*

- *A undercuts B (on B') iff $Conc(A) = -\mathtt{n}(r)$ for some $B' \in Sub(B)$ with $TopRule(B') = r$ and $r \in \mathcal{R}_d$*
- *A restrictively rebuts B (on B') iff $Conc(A) = -Conc(B')$ for some $B' \in Sub(B)$ with $TopRule(B') \in \mathcal{R}_d$*
- *A unrestrictively rebuts B (on B') iff $Conc(A) = -Conc(B')$ for some $B' \in Sub(B)$ with B' being a defeasible argument*

To illustrate the difference between restricted rebut and unrestricted rebut, first consider the example of an argumentation system AS_1 with $\mathcal{R}_s = \emptyset$ and $\mathcal{R}_d = \{\Rightarrow a;\ a \Rightarrow b;\ \Rightarrow c;\ c \Rightarrow \neg b\}$. Here, the argument $(\Rightarrow a) \Rightarrow b$ restrictively and unrestrictively rebuts the argument $(\Rightarrow c) \Rightarrow \neg b$, and vice versa. In the argumentation system AS_2 with $\mathcal{R}_s = \{\rightarrow a;\ a \rightarrow b\}$ and $\mathcal{R}_d = \{\Rightarrow c;\ c \Rightarrow \neg b\}$, the argument $(\rightarrow a) \rightarrow b$ restrictively and unrestrictively rebuts the argument $(\Rightarrow c) \Rightarrow \neg b$, but the argument $(\Rightarrow c) \Rightarrow \neg b$ does not restrictively or unrestrictively rebut the argument $(\rightarrow a) \rightarrow b$. In the argumentation system AS_3 with $\mathcal{R}_s = \{a \rightarrow b;\ \rightarrow c\}$ and $\mathcal{R}_d = \{\Rightarrow a;\ c \Rightarrow \neg b\}$ the argument $(\Rightarrow a) \rightarrow b$ restrictively and unrestrictively rebuts the argument $(\rightarrow c) \Rightarrow \neg b$, and the argument $(\rightarrow c) \Rightarrow \neg b$ unrestrictively (but not restrictively) rebuts the argument $(\Rightarrow a) \rightarrow b$. To sum up, with restrictive rebut one needs to check whether the *last* rule of the attacked conclusion[6] is defeasible whereas with unrestricted rebut one needs to check whether *any previous* rule of the attacked conclusion is defeasible.

The intuition behind unrestricted rebut is that a conclusion is defeasible iff it has been derived using at least one defeasible rule. If the conclusion has been derived using strict rules only, then the conclusion is strict and cannot be argued against. The intuition behind restricted rebut, on the other hand, is that (like in classical logic) in order to argue against a particular derivation, one has to argue against its premises. So instead of attacking the consequent of a strict rule, one has to attack its antecedent, unless this antecedent itself consists of the consequents of strict rules, in which case one has to keep on going backwards until finding a defeasible rule. It holds that if A restrictively rebuts B, then A also unrestrictively rebuts B, but not vice versa.

One last subtle aspect of the definition of restricted and unrestricted rebut (Definition 4) is that one only looks at the subargument B' that yields the conclusion that one is arguing against. So in the argumentation system AS_4 with $\mathcal{R}_s = \{\rightarrow c;\ c \rightarrow \neg b\}$ and $\mathcal{R}_d = \{\Rightarrow a;\ a \Rightarrow b;\ \neg b \Rightarrow d\}$ the argument $(\Rightarrow a) \Rightarrow b$

[6] meaning: of the conclusion one argues against by providing an argument for its contrary

does *not* (restrictively or unrestrictively) rebut the argument $((\rightarrow c) \rightarrow \neg b) \Rightarrow d$, even though the latter argument is defeasible, because the subargument that yields the attacked conclusion $\neg b$ is strict.

The difference between restricted and unrestricted rebut is relevant not just because they are based on different intuitions, but also because choosing to implement either restricted or unrestricted rebut has consequences for how one should define the rest of the argumentation formalism if the aim is to yield some kind of reasonable output in terms of justified conclusions. Details will follow further on in the current paper.

Apart from (restrictive and unrestrictive) rebutting, Definition 4 also introduces the concept of undercutting. Whereas with rebutting, one argues against the conclusion of an argument (or against the conclusion of a subargument), with undercutting one argues against the applicability of a particular defeasible rule. A classical example of undercutting has been given by Pollock [1995]: "If an object looks red, then it actually is red, unless it is illuminated by a red light". Formally, this can be modelled using argumentation system AS_5 with $\mathcal{R}_s = \{\rightarrow \mathit{looksred};\ \rightarrow \mathit{redlight}\}$, $\mathcal{R}_d = \{\mathit{looksred} \Rightarrow \mathit{isred};\ \mathit{redlight} \Rightarrow \neg \mathit{lris}\}$ and $\mathtt{n}(\mathit{looksred} \Rightarrow \mathit{isred}) = \mathit{lris}$. Here, the argument $(\rightarrow \mathit{looksred}) \Rightarrow \mathit{isred}$ is undercut by the argument $(\rightarrow \mathit{redlight}) \Rightarrow \neg \mathit{lris}$. Although undercutting does not play a major role in the remaining part of the current paper, we have still chosen to introduce it, as it is a piece of functionality that can be implemented while still warranting an overall reasonable outcome regarding the justified conclusions.

Another piece of functionality that some formalisms have implemented is that of argument strength.[7] Argument strength is often defined based on an ordering of the defeasible rules. However, as arguments can be constructed using more than one defeasible rule, one needs a way of applying the strength ordering between *individual* rules to determine a strength ordering between *sets* of rules. Two principles for doing so have been defined in the literature: the elitist and the democratic set ordering [Modgil and Prakken, 2014; Caminada *et al.*, 2014b].

Definition 5. *Let $\leq \subseteq (\mathcal{R}_d \times \mathcal{R}_d)$ be a total pre-ordering on the defeasible inference rules, where as usual, $r < r'$ iff $r \leq r'$ and $r \not\leq r'$, and $r \equiv r'$ iff $r \leq r'$ and $r' \leq r$. Then for any $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{R}_d$ $\trianglelefteq_{\mathtt{s}}$ ($\mathtt{s} \in \{\mathtt{Eli}, \mathtt{Dem}\}$) is defined as follows:*

1. *If $\mathcal{E} = \emptyset$ then $\mathcal{E} \not\trianglelefteq_{\mathtt{s}} \mathcal{E}'$;*
2. *If $\mathcal{E}' = \emptyset$ and $\mathcal{E} \neq \emptyset$ then $\mathcal{E} \trianglelefteq_{\mathtt{s}} \mathcal{E}'$; else:*

[7] Argument strength is sometimes referred to as *argument preferences* in the work of Prakken [2010], Modgil and Prakken [2014] and of Caminada *et al.* [2014b].

3. *if* $\mathtt{s} = \mathtt{Eli}$*:* $\mathcal{E} \trianglelefteq_{\mathtt{Eli}} \mathcal{E}'$ *if* $\exists r_1 \in \mathcal{E}$ *s.t.* $\forall r_2 \in \mathcal{E}'$, $r_1 \leq r_2$*; else:*

4. *if* $\mathtt{s} = \mathtt{Dem}$*:* $\mathcal{E} \trianglelefteq_{\mathtt{Dem}} \mathcal{E}'$ *if* $\forall r_1 \in \mathcal{E}$, $\exists r_2 \in \mathcal{E}'$, $r_1 \leq r_2$.

As usual $\mathcal{E} \triangleleft_{\mathtt{s}} \mathcal{E}'$ *iff* $\mathcal{E} \trianglelefteq_{\mathtt{s}} \mathcal{E}'$ *and* $\mathcal{E}' \not\trianglelefteq_{\mathtt{s}} \mathcal{E}$

The elitist and democratic set ordering principles assume the presence of sets of defeasible rules. This leads to the question of how to determine the relevant sets of defeasible rules when one argument rebuts another. Again, two principles have been formulated in the literature, called *weakest link* and *last link*. With weakest link, one takes into account *all* defeasible rules (of both the rebutting argument and the rebutted (sub)argument), whereas with last link, one takes into account only the *last* defeasible rule(s). Given the weakest link and the last link principles for determining the sets of relevant defeasible rules, as well as the elitist and democratic set ordering principles for evaluating these sets of defeasible rules, one can identify four different principles for determining argument strength.

Definition 6. *Let Ar be the set of arguments that can be constructed using argumentation system* $(\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$*. Then* $\forall A, B \in Ar$*:*

1. $A \preceq_{\mathtt{Ewl}} B$ *iff* $\mathtt{DefRules}(A) \trianglelefteq_{\mathtt{Eli}} \mathtt{DefRules}(B)$

2. $A \preceq_{\mathtt{Ell}} B$ *iff* $\mathtt{LastDefRules}(A) \trianglelefteq_{\mathtt{Eli}} \mathtt{LastDefRules}(B)$

3. $A \preceq_{\mathtt{Dwl}} B$ *iff* $\mathtt{DefRules}(A) \trianglelefteq_{\mathtt{Dem}} \mathtt{DefRules}(B)$

4. $A \preceq_{\mathtt{Dll}} B$ *iff* $\mathtt{LastDefRules}(A) \trianglelefteq_{\mathtt{Dem}} \mathtt{LastDefRules}(B)$

where Ewl*,* Ell*,* Dwl *and* Dll *respectively denote '*Elitist *weakest link', '*Elitist *last link', '*Democratic *weakest link' and '*Democratic *last link'.*
We may write $A \prec_p B$ *iff* $A \preceq_p B$ *and* $B \not\preceq_p A$*, and write* $A \approx_p B$ *iff* $A \preceq_p B, B \preceq_p A$ *(where* $p \in \{\mathtt{Ewl}, \mathtt{Ell}, \mathtt{Dwl}, \mathtt{Dll}\}$*). It is straightforward to show that* $\prec_p$ *is a strict partial ordering (irreflexive, transitive and asymmetric).*

We are now ready to define the overall notion of defeat. For this, we follow the approach of formalisms like ASPIC+ [Modgil and Prakken, 2014] and ASPIC- [Caminada *et al.*, 2014b], where the notion of defeat stands for attack after argument strength has been taken into account. It is defeat, not attack, that is then used to define the argumentation framework.

Definition 7. *Let Ar be the set of arguments that can be constructed using argumentation system* $AS = (\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$*. Let* $\preceq_p$ *be the associated argument strength order on Ar as defined in Definition 6. Then* $def_{ur} \subseteq Ar \times Ar$ *is defined as*

$(A, B) \in def_{ur}$ iff A undercuts B or A unrestrictively rebuts B on B' and $A \not\prec_p B'$, and $def_{rr} \subseteq Ar \times Ar$ is defined as $(A, B) \in def_{rr}$ iff A undercuts B or A restrictively rebuts B on B' and $A \not\prec_p B'$.

We observe that the set of arguments Ar, together with the associated defeat relation (either def_{ur} or def_{rr}) defines a Dung-style argumentation framework. On this argumentation framework, one can then apply the standard argumentation semantics.

3 Direct Consistency, Indirect Consistency and Closure

To illustrate the issue of rationality postulates, consider the following example.

Example 1 ([Caminada and Amgoud, 2007])**.** *Consider an argumentation system $AS = (\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ with $\mathcal{R}_s = \{\rightarrow r;\ \rightarrow n;\ m \rightarrow hs;\ b \rightarrow \neg hs\}$, $\mathcal{R}_d = \{r \Rightarrow m;\ n \Rightarrow b\}$, $\mathtt{n} = \emptyset$ and $\leq = \emptyset$.*

An intuitive interpretation of this example is the following:
"John wears a ring (r) on his finger. John is also a regular nightclubber (n). Someone who wears a ring on his finger is usually married (m). Someone who is a regular nightclubber is usually bachelor (b). Someone who's married by definition has a spouse (hs). Someone who's bachelor by definition does not have a spouse ($\neg hs$)."

We can construct the following arguments.

$A_1 : \rightarrow r$ $\quad A_3 : A_1 \Rightarrow m$ $\quad A_5 : A_3 \rightarrow hs$
$A_2 : \rightarrow n$ $\quad A_4 : A_2 \Rightarrow b$ $\quad A_6 : A_4 \rightarrow \neg hs$

If one were to apply unrestricted rebut, the only defeat would be between A_5 and A_6. That is, $def_{ur} = \{(A_5, A_6), (A_6, A_5)\}$. This then implies that for instance the grounded extension is $\{A_1, A_2, A_3, A_4\}$, yielding the associated set of (grounded) justified conclusions $\{r, n, m, b\}$. The problem with these conclusions, however, is that they do not take into account the meaning of the strict rules of the argumentation system: that if one holds the antecedent of a strict rule to be the case, one must also hold what deductively follows from it (the consequent of the rule). For instance, from the fact that we obtain m, together with the strict rule $m \rightarrow hs$ we should also have obtained hs, as a married person by definition has a spouse, so by John being married we cannot escape the conclusion that he has a spouse. Yet, the fact that John has a spouse is not represented in the set of justified conclusions (that is, $hs \notin \{r, n, m, b\}$). This brings us to the first problem: the set of justified conclusions is not closed under the strict rules.

Another problem appears when also applying the strict rule $b \rightarrow \neg hs$. After all, John is also considered to be a bachelor, so we cannot escape the conclusion that he

does not have a spouse ($\neg hs$). However, when we also apply the rule $m \rightarrow hs$, as we did earlier, then we derive that John both has a spouse and does not have a spouse. So not only is our set $\{r, n, m, b\}$ of justified conclusions not closed under the strict rules, if we do try to compute its closure, this closure turns out to be inconsistent!

So far, we examined what happens regarding the justified conclusions in case we apply unrestricted rebut. However, if we were to base the defeat relation on restricted rebut instead, then the outcome would even be worse, as the defeat relation would become empty (that is, $def_{rr} = \emptyset$) which means that (when still applying grounded semantics) one obtains $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ as the grounded extension and $\{r, n, m, b, hs, \neg hs\}$ as the associated justified conclusions. So here, we don't even need to close the justified conclusions under the strict rules in order to obtain an inconsistent outcome, as the set of justified conclusions is already inconsistent by itself.

From Example 1 we observe that there are at least three desirable properties a set of conclusions should satisfy.

Postulate 1. *Let $S \subseteq \mathcal{L}$ be a set of justified conclusions yielded by an argumentation system. S should satisfy:*

- direct consistency*, meaning that* $\neg\exists x : x, -x \in S$
- closure*, meaning that* $Cl_{\mathcal{R}_s}(S) = S$
- indirect consistency*, meaning that* $\neg\exists x : x, -x \in Cl_{\mathcal{R}_s}(S)$

Early formalisations of argumentation theory tried to avoid problems like those illustrated in Example 1 by tinkering with the definition of defeat. However, as explained by Caminada and Amgoud [2007], this does not actually lead to the properties of Postulate 1 being satisfied. Clearly, some more fundamental solutions are needed. In the following two subsections, we examine some of the solutions that have been described in the literature, distinguishing between solutions that have been obtained for restricted rebut and solutions that have been obtained for unrestricted rebut.

3.1 Restricted Rebut Solutions

In the current section, we examine some of the solutions that have been described in the literature for satisfying direct consistency, indirect consistency and closure when the defeat relation is based on restricted rebut.

We recall that, when applying restricted rebut to Example 1 this results in the empty defeat relation, that is $def_{rr} = \emptyset$. One could argue that this is because

something is wrong with the information encoded in the argumentation system AS, in particular with the set of strict rule $\mathcal{R}_s$. If one were for instance to add the additional strict rules $\neg hs \rightarrow \neg m$ and $hs \rightarrow \neg b$ then the problem would be solved. This is because one could then construct additional arguments $A_7 : \ A_5 \rightarrow \neg b$ and $A_8 : \ A_6 \rightarrow \neg m$. It holds that A_7 restrictively rebuts A_4 (as well as each argument that contains A_4, so also A_6 and A_8) and that A_8 restrictively rebuts A_3 (as well as each argument that contains A_3, so also A_5 and A_7). So overall we obtain the argumentation framework shown in Figure 1. This argumentation framework yields the grounded extension $\{A_1, A_2\}$ (with associated conclusions $\{r, n\}$) and preferred extensions $\{A_1, A_2, A_3, A_5, A_7\}$ (with associated conclusions $\{r, n, m, hs, \neg b\}$) and $\{A_1, A_2, A_4, A_6, A_8\}$ (with associated conclusions $\{r, n, b, \neg hs, \neg m\}$). As we can see, each set of conclusions yielded under grounded or preferred semantics satisfies the postulates of direct consistency, closure and indirect consistency.

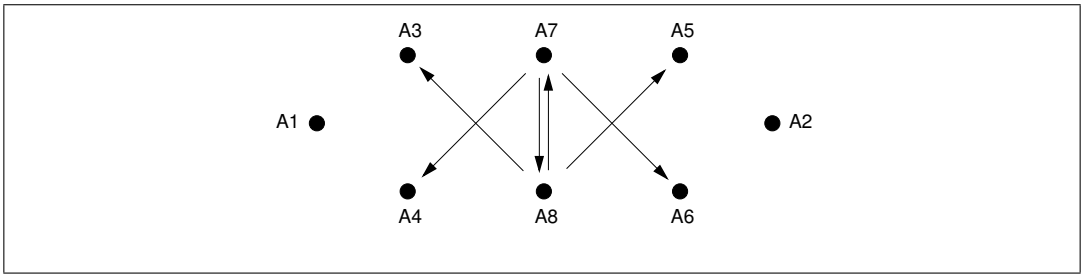

Figure 1: Argumentation framework of Example 1 after adding the rules $\neg hs \rightarrow \neg m$ and $hs \rightarrow \neg b$.

Adding the rules $\neg hs \rightarrow \neg m$ and $hs \rightarrow \neg b$ can be seen as a reasonable thing to do. After all, $\mathcal{R}_s$ already contains a rule $m \rightarrow hs$, meaning that without possible exception, someone who is married by definition has a spouse. This implies that someone who does not have a spouse cannot be married. Hence, $\neg hs \rightarrow \neg m$. Using similar reasoning, one can use the rule $b \rightarrow \neg hs$ to derive $hs \rightarrow \neg b$. Hence, the rules $\neg hs \rightarrow \neg m$ and $hs \rightarrow \neg b$ were already "implicitly" contained in $\mathcal{R}_s$. Adding them explicitly can therefore be seen as doing justice to $\mathcal{R}_s$, and has as a side effect that the postulates of direct consistency, closure and indirect consistency become satisfied.

Adding the "contraposed" version of a strict rule is relatively straightforward when the antecedent of the rule consists just of a single formula (as is for instance the case for $m \rightarrow hs$ and $b \rightarrow \neg hs$) but gets more complicated when the antecedent consists of multiple formulas. For this, a generalised version of contraposition is needed, which is referred to as *transposition* [Caminada and Amgoud, 2007].

Definition 8 ([Caminada and Amgoud, 2007]). *Let $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ ($n \geq 0$) be a strict rule. A transposed version of this rule is of the form $\varphi_1, \ldots, \varphi_{i-1}, -\varphi, \varphi_{i+1}, \ldots, \varphi_n \rightarrow -\varphi$ (for some $i \in \{1 \ldots n\}$). We say that a set of strict rules $\mathcal{R}_s$ is closed under transposition when for each strict rule in $\mathcal{R}_s$, each of its transposed versions is also in $\mathcal{R}_s$.*

As an example, the strict rule $a, \neg b, c \rightarrow d$ has three transposed versions: $\neg d, \neg b, c \rightarrow \neg a$; $a, \neg d, c \rightarrow b$ and $a, \neg b, \neg d \rightarrow \neg c$.

An example of an argumentation formalism that applies transposition to satisfy direct consistency, closure and indirect consistency is ASPIC+ [Modgil and Prakken, 2014]. In ASPIC+ the following design choices have been made:

- the set of strict rules $\mathcal{R}_s$ is consistent and closed under transposition
- restricted rebut is applied
- argument strength is based on a partial pre-order on the defeasible rules, together with either the last-link or weakest link selection principle and either the elitist or democratic set ordering principle[8]
- the argumentation semantics is complete-based, meaning that it selects one or more complete extensions (examples of complete-based semantics are grounded, preferred, complete, semi-stable, ideal and eager semantics)

It is shown that under these choices, the overall outcome of the formalism satisfies direct consistency, closure and indirect consistency.

To understand why transposition plays an important role in satisfying the properties of direct consistency, closure and indirect consistency, it can be useful to give a sketch of proof. We start with the property of direct consistency. Suppose, towards a contradiction, that there exists a complete extension yielding conclusions that are directly inconsistent. This means there exists an argument A for conclusion c and an argument B for conclusion $-c$ (see Figure 2). As the set of strict rules $\mathcal{R}_s$ is consistent, at least one of these arguments must be defeasible. Assume without loss of generality that argument A is defeasible. Then A must contain at least one defeasible rule. Now, identify a defeasible rule r that is "as high as possible" in A (that is, whose distance to the conclusion c is minimal). Let e be the consequent of r and let A_i be the subargument of A that has r as its top rule (so $Conc(A_i) = e$). Let $A_1, \ldots, A_n$ be the subarguments of A that have the same "depth" as A_i (that

[8] More precisely, argument strength has to be based on a *reasonable argument ordering* [Modgil and Prakken, 2014], which is satisfied by applying either the weakest link or the last link selection principle, in combination with applying either the democratic or the elitist set ordering principle.

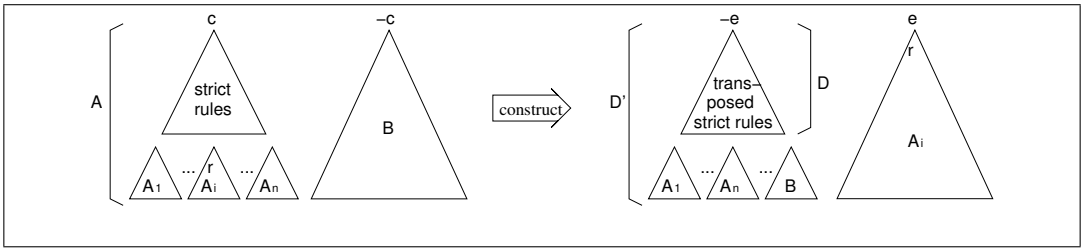

Figure 2: Sketch of proof direct consistency (restricted rebut)

is, whose respective top-rules have the same distance to conclusion c). It turns out to be possible to build an argument D' that defeats A_i by deriving conclusion $-e$. Recall that "above" each A_i there are only strict rules in A (after all, r was the "highest" defeasible rule in A). In case these strict rules consist of only one layer, there exists a single strict rule $Conc(A_1), \ldots, Conc(A_n) \rightarrow c$) with transposed version $Conc(A_1), \ldots, Conc(A_{i-1}), -c, Conc(A_{i+1}), \ldots, Conc(A_n) \rightarrow -Conc(A_i)$, so $Conc(A_1), \ldots, Conc(A_{i-1}), Conc(B), Conc(A_{i+1}), \ldots, Conc(A_n) \rightarrow -c$, which implies we can use $A_1, \ldots A_{i-1}$, B and $A_{i+1}, \ldots, A_n$ to construct an argument that restrictively rebuts A_i. In case the strict rules above each A_i consist of more than one layer, then one can still use transposition to construct an argument that restrictively rebuts A_i (basically by induction over the number of layers of strict rules). Let D' be the thus constructed argument that restrictively rebuts A_i. As A_i is a subargument of A, it follows that D' also restrictively rebuts A. From the fact that we are considering a complete extension, it follows that the extension has to contain an argument (say C) that defeats D'. However, as each defeasible rule of D' also occurs in A or B, it follows that C also defeats A or B.[9] Hence, the complete extension is not conflict-free. Contradiction.

It is important to observe that the above sketch of proof uses the facts that (1) $\mathcal{R}_s$ is consistent, (2) $\mathcal{R}_s$ is closed under transposition, (3) restricted rebut is being applied, and (4) we are considering a complete extension (or at least an admissible set).[10]

As for the property of closure, suppose there exists a strict rule $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and that the conclusions $\varphi_1, \ldots, \varphi_n$ are yielded by our complete extension. We need to show that conclusion φ is also yielded by the complete extension. From the fact

[9] This is straightforward to see when the strength ordering between the rules is empty, but also holds when the strength ordering is non-empty. See the work of Modgil and Prakken [2013] for details.

[10] There are also some requirements regarding argument strength. These are such that $\preceq_{\texttt{Ewl}}$, $\preceq_{\texttt{Ell}}$, $\preceq_{\texttt{Dwl}}$, and $\preceq_{\texttt{Dll}}$ (Definition 6) satisfy them. We refer to the work of Modgil and Prakken [2013; 2014] for details.

that conclusions $\varphi_1, \ldots, \varphi_n$ are yielded, it follows that the complete extension contains arguments $A_1, \ldots, A_n$ with conclusions $\varphi_1, \ldots, \varphi_n$ respectively. Now consider the argument $A : A_1, \ldots, A_n \rightarrow \varphi$. Let B be an arbitrary argument that defeats A. Then from the definition of defeat, it follows that B also defeats at least one of $A_1, \ldots, A_n$. From the fact that our extension is complete (and therefore also admissible) it follows that it contains an argument (say C) that defeats B. This means that A is defended by the complete extension, and must therefore also be contained in the complete extension.[11] This then implies that the complete extension also yields conclusion $Conc(A) = \varphi$.

Given that we have obtained both direct consistency and closure, the property of indirect consistency is trivially satisfied.

As was mentioned above, the property of transposition plays an important role for satisfying direct consistency, closure and indirect consistency. However, if one takes a closer look at the above sketch of proof, what is actually applied is a property that is more general than transposition. Going back to Figure 2 then what is actually needed is that if from $Conc(A_1), \ldots, Conc(A_n)$ one can apply strict rules to derive c, then from $Conc(A_1), \ldots, Conc(A_{i-1}), -c, Conc(A_{i+1}), \ldots, Conc(A_n)$ one can also apply strict rules to derive $-Conc(A_i)$. This property is called *contraposition* by Modgil and Prakken [2013; 2014], who show that direct consistency, closure and indirect consistency are satisfied when the set of strict rules is closed under contraposition.

One can ask the question of whether it is possible to derive even more general conditions than transposition and contraposition, under which direct consistency, closure and indirect consistency are still satisfied. This question is answered positively by Dung and Thang [2014] who present a semi-abstract approach that abstracts away from most aspects of argument structure (making explicit only the notions of a conclusion and that of a subargument). However, their approach does rely on particular constraints on the defeat relation, and it can be observed that these constraints can only be satisfied under restricted (and not unrestricted) rebut.[12]

3.2 Unrestricted Rebut Solutions

Although restricted rebut has become the most popular principle for defining the overall defeat relationship (as is for instance evidenced by the various versions of the

[11] Notice that for this reasoning step, a complete extension is really needed; an admissible set is not sufficient.

[12] More precisely, unrestricted rebut trivialises the notion of a *base* [Dung and Thang, 2014], which prevents the results of Dung and Thang [2014] from being applied in the context of unrestricted rebut.

ASPIC+ formalism [Prakken, 2010; Modgil and Prakken, 2013; Modgil and Prakken, 2014]) it does have some disadvantages, especially when applied in a dialectical context. Consider for instance the following discussion taken from [Caminada *et al.*, 2014b].

John: *"Bob will attend both AAMAS and IJCAI this year, as he has papers accepted at each of these conferences."*

Mary: *"That won't be possible, as his budget of £1000 only allows for one foreign trip."*

Formally, this discussion can be modelled using the argumentation system $(\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ with $\mathcal{R}_d = \{accA \Rightarrow attA;\ accI \Rightarrow attI;\ budget \Rightarrow \neg attboth\}$ and $\mathcal{R}_s = \{\rightarrow accA;\ \rightarrow accI;\ \rightarrow budget;\ attA, attI \rightarrow attboth;\ \neg attboth, attI \rightarrow \neg attA;\ attA, \neg attboth \rightarrow \neg attI\}$.[13]

John: $((\rightarrow accA) \Rightarrow attA), ((\rightarrow accI) \Rightarrow attI) \rightarrow attboth$

Mary: $(\rightarrow budget) \Rightarrow \neg attboth$

The problem is that when applying restricted rebut, Mary's argument does not defeat John's argument. This is because the conclusion that Mary wants to attack ($attboth$) is the consequent of a strict rule. If Mary wants to restrictively rebut John's argument, she can only do so by attacking the consequent of a defeasible rule. That is, she would be forced to choose to defeat either $attA$ or $attI$, meaning that she essentially has to utter one of the following statements.

Mary$'$: *Bob won't attend AAMAS because he will already attend IJCAI, and his budget doesn't allow him to attend both.*

Mary$''$: *Bob won't attend IJCAI because he will already attend AAMAS, and his budget doesn't allow him to attend both.*

The associated formal counterarguments are as follows.

Mary$'$: $((\rightarrow budget) \Rightarrow \neg attboth), ((\rightarrow accI) \Rightarrow attI) \rightarrow \neg attA$

Mary$''$: $((\rightarrow accA) \Rightarrow attA), ((\rightarrow budget) \Rightarrow \neg attboth) \rightarrow \neg attI$

Critically, Mary does not *know* which of the two conferences Bob will attend, yet the principle of restricted rebut *forces* her to make concrete statements on this. From the perspective of commitment in dialogue [Walton and Krabbe, 1995], this is unnatural. One should not be forced to commit to things one has insufficient reasons to believe in.

It should be stressed that the problem outlined above is particularly relevant in dialectical contexts, where different agents make commitments during the exchange of arguments. This contrasts with a formalism like ASPIC+, which is more monolithic in nature, in that from the given rules and premises, one constructs a graph of each other defeating arguments and simply *computes* which arguments (and associated

[13] We observe that $\mathcal{R}_s$ is consistent and closed under transposition.

conclusions) are justified. Concepts like different agents, communication steps or commitment stores do not play a role in ASPIC+, and hence restricted rebut *seems* acceptable. However, if one wants to add dialectical aspects to formal argumentation (c.f., [Caminada and Wu, 2009; Caminada and Podlaszewski, 2012; Caminada *et al.*, 2014a]) then one is forced to take the limitations of restricted rebut seriously.

The obvious way to deal with problems like sketched above would be to simply replace restricted rebut by unrestricted rebut (thus replacing def_{rr} by def_{ur}). Unfortunately, doing so also has far reaching consequences regarding the ability to satisfy the postulates of indirect consistency and closure. This is illustrated by the following example, taken from [Caminada and Wu, 2011].

Example 2. *Consider the argumentation system* $(\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ *with* $\mathcal{R}_s = \{\rightarrow jw;\ \rightarrow mw;\ \rightarrow sw;\ mt, st \rightarrow \neg jt;\ jt, st \rightarrow \neg mt;\ jt, mt \rightarrow \neg st\}$ *and* $\mathcal{R}_d = \{jw \Rightarrow jt;\ mw \Rightarrow mt;\ sw \Rightarrow st\}$*. This example can be interpreted as follows. John, Mary and Suzy want to go cycling in the countryside* ($\rightarrow jw;\ \rightarrow mw;\ \rightarrow sw$)*. They have a tandem bicycle that each of them would like to be on* ($jw \Rightarrow jt;\ mw \Rightarrow mt;\ sw \Rightarrow st$)*. However, as the tandem only has two seats, if two of them are on it, the third one cannot be on it* ($mt, st \rightarrow \neg jt;\ jt, st \rightarrow \neg mt;\ jt, mt \rightarrow \neg st$)*. Using this argumentation system, we can construct the following arguments.*

$A_1 : \rightarrow jw$ $\quad A_4 : A_1 \Rightarrow jt$ $\quad A_7 : A_5, A_6 \rightarrow \neg jt$
$A_2 : \rightarrow mw$ $\quad A_5 : A_2 \Rightarrow mt$ $\quad A_8 : A_4, A_6 \rightarrow \neg mt$
$A_3 : \rightarrow sw$ $\quad A_6 : A_3 \Rightarrow st$ $\quad A_9 : A_4, A_5 \rightarrow \neg st$

When applying restricted rebut (and assuming the empty rule strength ordering) argument A_7 *defeats* A_4 *(as well as* A_8 *and* A_9*, which contain* A_4*), argument* A_8 *defeats* A_5 *(as well as* A_7 *and* A_9*, which contain* A_5*) and argument* A_9 *defeats* A_6 *(as well as* A_7 *and* A_8*, which contain* A_6*). This yields the argumentation framework at the left hand side of Figure 3, which we will refer to as* AF_{rr}*.*

AF_{rr} *has four complete extensions:* $\{A_1, A_2, A_3, A_5, A_6, A_7\}$ *(yielding conclusions* $\{jw, mw, sw, \neg jt, mt, st\}$*),* $\{A_1, A_2, A_3, A_4, A_6, A_8\}$ *(yielding conclusions* $\{jw, mw, sw, jt, \neg mt, st\}$*),* $\{A_1, A_2, A_3, A_4, A_5, A_9\}$ *(yielding conclusions* $\{jw, mw, sw, jt, mt, \neg st\}$*), and* $\{A_1, A_2, A_3\}$ *(yielding conclusions* $\{jw, mw, sw\}$*). The first three complete extensions are also preferred (as well as stable and semi-stable). The last one is also grounded. We observe that the conclusions of each complete extension satisfy direct consistency, closure and indirectly consistency.*

Now, let us consider what happens if we were to replace restricted rebut by unrestricted rebut. In that case, A_7 *would still defeat* A_4 *(as well as* A_8 *and* A_9*),* A_8 *would still defeat* A_5 *(as well as* A_7 *and* A_9*) and* A_9 *would still defeat* A_6 *(as well as* A_7 *and* A_8*). However, additionally* A_4 *would defeat* A_7*,* A_5 *would defeat* A_8 *and* A_6 *would defeat* A_9*. This is because* A_7*,* A_8 *and* A_9 *are defeasible arguments, as*

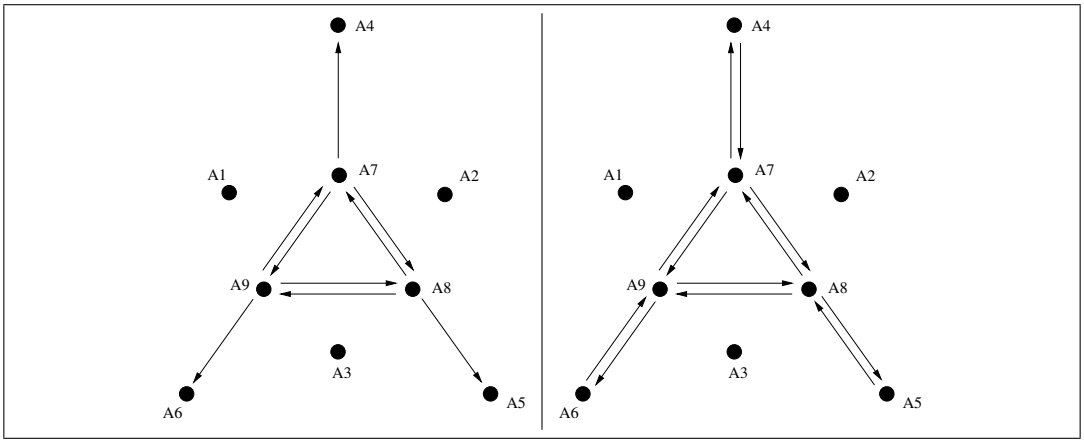

Figure 3: restricted rebut versus unrestricted rebut

their subarguments contain defeasible rules. So with unrestricted rebut, the arguments A_4, A_5 and A_6 are able to "strike back" against their respective defeaters. This yields the argumentation framework at the right hand side of Figure 3, which we will refer to as AF_{ur}. AF_{ur} has five complete extensions. The first four are the same as those of AF_{rr}. The fifth one is $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ yielding conclusions $\{jw, mw, sw, jt, mt, st\}$, hence violating closure and indirect consistency. As this fifth complete extension is also preferred, stable and semi-stable, we have a counterexample against applying unrestricted rebut under each of these semantics.

Example 2 illustrates a fundamental difference between restricted and unrestricted rebut. Whereas under restricted rebut (in combination with $\mathcal{R}_s$ being consistent and closed under transposition or contraposition) any admissible set of arguments will yield conclusions that are indirectly consistent, under unrestricted rebut admissibility alone is not sufficient (the set $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ being the counter example). It turns out that what is needed is a property that is stronger than admissibility: strong admissibility [Baroni and Giacomin, 2009; Caminada, 2014].[14] We observe that although the set $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ is admissible, it is not strongly admissible. Furthermore, we observe that the set $\{A_1, A_2, A_3\}$ is both admissible and strongly admissible and yields conclusions $\{jw, mw, sw\}$ that are closed and indirectly consistent.

As the grounded extension is the unique biggest strongly admissible set [Baroni and Giacomin, 2009; Caminada, 2014], grounded semantics is a natural starting

[14] We recall that a set of arguments $\mathcal{A}rgs$ is strongly admissible iff each $A \in \mathcal{A}rgs$ is defended by some $\mathcal{A}rgs' \subseteq \mathcal{A}rgs \setminus \{A\}$ which in its turn is again strongly admissible. Informally, the idea of strong admissibility is that each argument should be defended without going around in circles.

point for proving the properties of direct consistency, indirect consistency and closure when applying unrestricted rebut. Proving the property of direct consistency is relatively straightforward. After all, if the grounded extension was to yield conclusions that are directly inconsistent, it would have to contain two arguments A and B with opposite conclusions. As $\mathcal{R}_s$ is consistent, at least one of them has to be defeasible, which means that one would defeat (unrestrictedly rebut) the other, which would implies that the grounded extension is not conflict-free. Contradiction.

Proving the property of closure is a bit more complex, as it is done by induction using the inductive definition of the grounded extension. We refer to the work of Caminada and Amgoud [2007] and of Caminada *et al.* [2014b] for details. Indirect consistency then follows trivially from direct consistency and closure.

As for argument strength, two possibilities have been observed when it comes to satisfying closure and indirect consistency under unrestricted rebut. The first approach, of Caminada and Amgoud [2007], is to essentially have the empty ordering on the defeasible rules. A later approach, by Caminada *et al.* [2014b] is to have a total (!) pre-order among the defeasible rules.

An overall overview of approaches to satisfy direct consistency, closure and indirect consistency is provided in Table 1.

4 Non-Interference and Crash Resistance

One of the issues to decide when formulating an argumentation system is whether the (strict and defeasible) rules should be domain dependent or domain independent. An example of a domain dependent strict rule would be $cow \rightarrow mammal$. An example of a domain independent strict rule would be modus ponens, so cow, $cow \supset mammal \rightarrow mammal$. When the aim is to implement domain independent reasoning, the most obvious thing to do would be to base the strict rules on some form of classical logic. For current purposes, we examine what happens if one were to base the set of strict rules on propositional logic.

Definition 9. *Given the language $\mathcal{L}$ of propositional logic, a* defeasible theory *is a tuple $(P, \mathcal{R}_d, \mathtt{n}, \leq)$ where*

- *P is a consistent set of propositions (called* premises*)*
- *$\mathcal{R}_d$ is a set of defeasible rules of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ (where φ_i, φ are meta-variables ranging over $\mathcal{L}$)*
- *$\mathtt{n}$ is a function such that $\mathtt{n} : \mathcal{R}_d \longrightarrow \mathcal{L}$*

defeat based on	argument strength	semantics	other conditions	example formalism
restricted rebut	empty	any complete-based semantics	$\mathcal{R}_s$ consistent and closed under transposition	ASPIC [Caminada and Amgoud, 2007]
unrestricted rebut	empty	grounded semantics	$\mathcal{R}_s$ consistent and closed under transposition	ASPIC [Caminada and Amgoud, 2007]
restricted rebut	partial pre-order $\mathcal{R}_d$, last link or weakest link, elitist or democratic	any complete-based semantics	$\mathcal{R}_s$ consistent and closed under transposition/ contraposition	ASPIC+ [Modgil and Prakken, 2014]
unrestricted rebut	total pre-order $\mathcal{R}_d$, last link or weakest link, elitist or democratic	grounded semantics	$\mathcal{R}_s$ consistent and closed under transposition	ASPIC− [Caminada *et al.*, 2014b]

Table 1: Approaches for satisfying closure and direct/indirect consistency

Given a defeasible theory $(P, \mathcal{R}_d, \mathtt{n}, \leq)$*, we define the associated argumentation system as* $(\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ *with* $\mathcal{R}_s = \{\rightarrow \varphi \mid \varphi \in P\} \cup \{\varphi_1, \ldots, \varphi_n \rightarrow \varphi \mid \varphi_1, \ldots, \varphi_n \vdash \varphi\}$

As P is a consistent set of formulas, $\mathcal{R}_s$ will be consistent. Moreover, $\mathcal{R}_s$ is also closed under transposition. This is because the set $\{\rightarrow \varphi \mid \varphi \in P\}$ is trivially closed under transposition (as a rule with an empty antecedent does not have any transposed versions) and the set $\{\varphi_1, \ldots, \varphi_n \rightarrow \varphi \mid \varphi_1, \ldots, \varphi_n \vdash \varphi\}$ is closed under transposition as $\varphi_1, \ldots, \varphi_n \vdash \varphi$ implies $\varphi_1, \ldots, \varphi_{i-1}, -\varphi, \varphi_{i+1}, \ldots, \varphi_n \vdash -\varphi$. However, basing strict rules on classical logic also brings an additional type of problems. Consider the following example.

Example 3. *Consider the defeasible theory* $(P, \mathcal{R}_d, \mathtt{n}, \leq)$ *with* $P = \{js, mns\}$, $\mathcal{R}_d = \{js \Rightarrow s;\ mns \Rightarrow \neg s;\ wfr \Rightarrow r\}$ *and* $\mathtt{n}$ *and* $\leq$ *being the empty ordering. This example can be interpreted as follows. John says the cup of coffee contains sugar,*

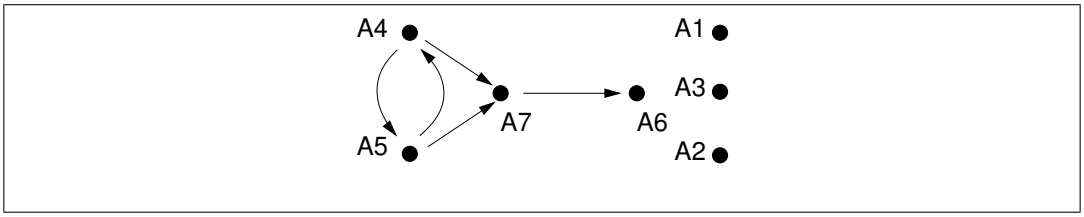

Figure 4: Strict rules as classical logic can have side effects (simple example)

so it probably contains sugar ($\rightarrow js$; $js \Rightarrow s$). *Mary says the cup of coffee does not contain sugar* ($\rightarrow mns$; $mns \Rightarrow \neg s$). *The weather forecaster predicts rain tomorrow, so it will rain tomorrow* ($\rightarrow wfr$; $wfr \Rightarrow r$). *Hence, although we're not sure about whether the cup of coffee contains sugar, at least we should believe that it will rain tomorrow. Using this argumentation system, at least the following arguments can be constructed.*

$A_1 : \rightarrow js$ $\quad A_4 : A_1 \Rightarrow s$

$A_2 : \rightarrow mns$ $\quad A_5 : A_2 \Rightarrow \neg s$

$A_3 : \rightarrow wfr$ $\quad A_6 : A_3 \Rightarrow r$

However, classical logic also yields the strict rule $s, \neg s \rightarrow \neg r$, *as* $s, \neg s \vdash \neg r$ (ex falso quodlibet). *With this rule, we can construct the following argument.*

$A_7 : A_4, A_5 \rightarrow \neg r$

This yields the argumentation framework of Figure 4.[15]

If one were to apply for instance grounded semantics, the grounded extension $\{A_1, A_2, A_3\}$ *would yield conclusions* $\{j, m, wf\}$. *Thus, the weather forecast is not believed because John and Mary are having a disagreement about a cup of coffee.*

The first thing to observe about Example 3 is that the underlying problem cannot be solved simply by removing rules with an inconsistent antecedent. This is because the effects of the rule $s, \neg s \rightarrow \neg r$ can be simulated by the rules $s \rightarrow s \vee \neg r$ and $s \vee \neg r, \neg s \rightarrow \neg r$, which still allow us to construct an argument for $\neg r$ from A_4 and A_5.

One approach that has been proposed in the literature [Prakken, 2010] is to change the semantics. If one were to apply for instance not grounded but preferred semantics to the argumentation framework of Figure 4, then two extensions would result: $\{A_1, A_2, A_3, A_4, A_6\}$ (yielding conclusions $\{j, m, wf, s, r\}$) and $\{A_1, A_2, A_3, A_5, A_6\}$ (yielding conclusions $\{j, m, wf, \neg s, r\}$). We observe that each set of conclusions contains r, so r is a justified conclusion under preferred semantics.

Although changing grounded semantics to preferred semantics seems to yield the

[15] Notice that we are applying restricted rebut, but similar problems also occur when applying unrestricted rebut.

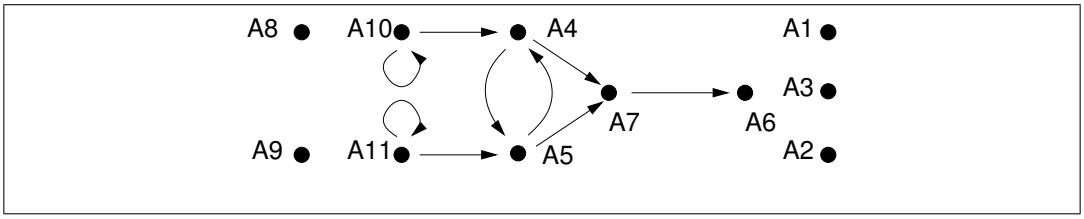

Figure 5: Strict rules as classical logic can have side effects (complex example)

desired outcome in Example 3, there exists a slightly more complex example where preferred semantics does *not* yield the desired outcome.

Example 4. *Consider the defeasible theory* $(P, \mathcal{R}_d, \mathtt{n}, \leq)$ *with* $P = \{js, mns, junrel, munrel, wfr\}$, $\mathcal{R}_d = \{js \Rightarrow s;\ mns \Rightarrow \neg s;\ wfr \Rightarrow r;\ junrel \Rightarrow \neg jrel;\ munrel \rightarrow \neg mrel\}$, $\mathtt{n}(js \Rightarrow s) = \mathtt{n}(junrel \Rightarrow \neg jrel) = jrel$, $\mathtt{n}(mns \Rightarrow \neg s) = \mathtt{n}(munrel \Rightarrow \neg mrel) = mrel$ *and* $\leq$ *being the empty ordering. So now, in addition to John saying that the cup of coffee contains sugar, he also says that he is unreliable, so John is probably unreliable* $(junrel \Rightarrow \neg jrel)$*. However, if John is unreliable, then the fact that he says something is no longer a reason to believe it. Hence the rule* $(js \Rightarrow s)$ *is undercut, just like the rule* $(junrel \Rightarrow \neg jrel)$*. Similarly, in addition to Mary saying that the cup of coffee does not contain sugar, she also says that she is unreliable, so Mary is probably unreliable* $(munrel \Rightarrow \neg mrel)$*. However, if Mary is unreliable, then the fact that she says something is no longer a reason to believe it. Hence the rule* $(mns \Rightarrow \neg s)$ *is undercut, just like the rule* $(munrel \Rightarrow \neg mrel)$*. Overall, we can construct at least the following arguments.*

$A_1 : \rightarrow js$ $\quad$ $A_4 :\ A_1 \Rightarrow s$
$A_2 : \rightarrow mns$ $\quad$ $A_5 :\ A_2 \Rightarrow \neg s$
$A_3 : \rightarrow wfr$ $\quad$ $A_6 :\ A_3 \Rightarrow r$
$A_8 : \rightarrow junrel$ $\quad$ $A_{10} :\ A_8 \Rightarrow \neg jrel$
$A_9 : \rightarrow munrel$ $\quad$ $A_{11} :\ A_9 \Rightarrow \neg mrel$

Classical logic again yields the strict rule $s, \neg s \rightarrow \neg r$*, which allows the construction of the following argument.*

$A_7 :\ A_4, A_5 \rightarrow \neg r$

This yields the argumentation framework of Figure 5.[16]

In the argumentation framework of Figure 5 there exists just a single complete extension (that is also grounded, preferred, ideal and semi-stable): $\{A_1, A_2, A_3, A_8, A_9\}$ *yielding conclusions* $\{js, mns, wfr, junrel, munrel\}$*. So again, we have that the weather forecast is not believed (under* any *admissibility-based semantics) because*

[16]Notice that we are again applying restricted rebut, although similar problems also occur when applying unrestricted rebut.

John and Mary are having a disagreement about a cup of coffee.

Before continuing to discuss some solutions that have been proposed in the literature, it can be useful to first define what precisely is it that we are trying to satisfy. Or, to put it in other words, what is the property that is actually being violated in Example 3 and Example 4? For this, we follow the approach of Caminada *et al.* [2012].

First of all, if $DT = (P, \mathcal{R}_d, \mathtt{n}, \leq)$ is a defeasible theory, then we write $\mathtt{Atoms}(DT)$ for the set of all propositional atoms occurring in DT. We say that defeasible theories DT_1 and DT_2 are syntactically disjoint iff $\mathtt{Atoms}(DT_1) \cap \mathtt{Atoms}(DT_2) = \emptyset$. For syntactically disjoint defeasible theories $DT_1 = (P_1, \mathcal{R}_{d1}, \mathtt{n}_1, \leq_1)$ and $DT_2 = (P_2, \mathcal{R}_{d2}, \mathtt{n}_2, \leq_2)$ we define the union $DT_1 \cup DT_2$ as $(P_1 \cup P_2, \mathcal{R}_{d1} \cup \mathcal{R}_{d2}, \mathtt{n}_1 \cup \mathtt{n}_2, \leq_1 \cup \leq_2)$. Also, given a defeasible theory DT, we define its consequences $Cn_\sigma(DT)$ as $\{Concs(\mathcal{A}rgs_1\}, \ldots, Concs(\mathcal{A}rgs_n)\}$ where $\mathcal{A}rgs_1, \ldots, \mathcal{A}rgs_n$ are the extensions of arguments (under semantics σ) of the argumentation framework yielded by defeasible theory DT. Given a set of propositions S and a set of propositional atoms $\mathcal{A}$, we define $S_{|\mathcal{A}}$ as $\{\varphi \in S \mid \text{each atom in } \varphi \text{ is an element of } \mathcal{A}\}$. Similarly, given a set $\mathcal{S} = \{S_1, \ldots, S_n\}$ where each S_i $(i \in \{1 \ldots n\})$ is a set of propositions, we define $\mathcal{S}_{|\mathcal{A}}$ as $\{S_{1|\mathcal{A}}, \ldots, S_{n|\mathcal{A}}\}$.

Definition 10. *An argumentation formalism (applying semantics σ) satisfies non-interference iff for every pair of syntactically disjoint defeasible theories DT_1 and DT_2 it holds that $Cn_\sigma(DT_1)_{|\mathtt{Atoms}(DT_1)} = Cn_\sigma(DT_1 \cup DT_2)_{|\mathtt{Atoms}(DT_1)}$.*

To see how non-interference can be violated, consider again Example 3. In essence, the defeasible theory of this example can be seen as the union of two syntactically disjoint defeasible theories $DT_1 = (P_1, \mathcal{R}_{d1}, \mathtt{n}_1, \leq_1)$ and $DT_2 = (P_2, \mathcal{R}_{d2}, \mathtt{n}_2, \leq_2)$ with $P_1 = \{wfr\}$, $\mathcal{R}_{d1} = \{wfr \Rightarrow r\}$, $P_2 = \{js, mns\}$, $\mathcal{R}_{d2} = \{js \Rightarrow s;\ mns \Rightarrow \neg s\}$, $\mathtt{n}_1 = \mathtt{n}_2 = \emptyset$ and $\leq_1 = \leq_2 = \emptyset$. When applying grounded semantics, it holds that $Cn_{gr}(DT_1)_{|\mathtt{Atoms}(DT_1)} = \{\{wfr, r\}\}$ whereas $Cn_{gr}(DT_1 \cup DT_2)_{|\mathtt{Atoms}(DT_1)} = \{\{wfr\}\}$. So merging DT_1 with the completely unrelated defeasible theory DT_2 affects the outcome that is relevant w.r.t. DT_1. Hence, non-interference is violated.

An even stronger property is that of crash resistance.

Definition 11. *A defeasible theory $DT_1 = (P_1, \mathcal{R}_{d1}, \mathtt{n}_1, \leq_1)$ (with $\mathtt{Atoms}(DT_1) \subsetneq \mathtt{Atoms}(\mathcal{L})$) is called contaminating (under semantics σ) iff for each syntactically disjoint defeasible theory DT_2 it holds that $Cn_\sigma(DT_1) = Cn_\sigma(DT_1 \cup DT_2)$. An argumentation formalism satisfies crash resistance iff there exists no defeasible theory that is contaminating.*

To see how crash resistance can be violated, consider Example 4. Again, the defeasible theory of this example can be seen as the union of two syntactically

disjoint defeasible theories $DT_1 = (P_1, \mathcal{R}_{d1}, \mathtt{n}_1, \leq_1)$ and $DT_2 = (P_2, \mathcal{R}_{d2}, \mathtt{n}_2, \leq_2)$ with $P_1 = \{js, mns, junrel, munrel\}$, $\mathcal{R}_{d1} = \{js \Rightarrow s;\ mns \Rightarrow \neg s;\ junrel \Rightarrow \neg jrel;\ munrel \Rightarrow \neg mrel\}$, $\mathtt{n}_1(js \Rightarrow s) = \mathtt{n}_1(junrel \Rightarrow \neg jrel) = jrel$, $\mathtt{n}_1(mns \Rightarrow \neg s) = \mathtt{n}_1(munrel \Rightarrow \neg mrel) = mrel$, $\leq_1 = \emptyset$, $P_2 = \{wfr\}$, $\mathcal{R}_{d2} = \{wfr \Rightarrow r\}$, $\mathtt{n}_2 = \emptyset$ and $\leq_2 = \emptyset$. When applying stable semantics, it holds that $Cn_{st}(DT_1) = \emptyset$, just like $Cn_{st}(DT_1 \cup DT_2) = \emptyset$. Moreover, it can be verified that for *any* DT_2' that is syntactically disjoint with DT_1, it holds that $Cn_{st}(DT_1 \cup DT_2') = \emptyset$, hence violating crash resistance under stable semantics.

Conceptually, the difference between non-interference and crash resistance is as follows. A violation of non-interference means that a defeasible theory somehow influences the entailment of a completely unrelated (syntactically disjoint) defeasible theory when being merged to it. A violation of crash resistance is more severe, as this means that a defeasible theory influences the entailment of a completely unrelated (syntactically disjoint) defeasible theory to such an extent that the actual contents of this other defeasible theory become totally irrelevant. An argumentation formalism that satisfies non-interference also satisfies crash resistance.[17]

Now that the relevant properties have been identified, we proceed to examine some of the approaches in the literature for satisfying these. The first approach to be discussed is that of Wu and Podlaszewski [2015]. Their main idea is simply to erase inconsistent arguments[18] from the argumentation framework before applying argumentation semantics.

Definition 12. *Let (Ar, def) be the argumentation framework constructed from defeasible theory DT (by applying restricted rebut). Let Ar_c be $\{A \in Ar \mid A \text{ is consistent}\}$ and let def_c be $def \cap (Ar_c \times Ar_c)$. (Ar_c, def_c) is defined as the inconsistency cleaned argumentation framework of DT.*

As an example of how Definition 12 is used, in Example 3 and Example 4 argument A_7 would be removed, as well as all attacks from and to A_7. The resulting inconsistency cleaned argumentation framework is such that r is a conclusion of each complete extension.

One of the main results proved by Wu and Podlaszewski [2015] is that removing inconsistent arguments from the argumentation framework does not lead to any violations of direct consistency, closure and indirect consistency.[19] They also prove

[17] That is, as long as the argumentation formalism is *non-trivial* in the sense of [Caminada *et al.*, 2012].

[18] An argument A is called inconsistent iff $\{Conc(A') \mid A' \in Sub(A)\}$ is inconsistent.

[19] This is unlike what for instance would happen when removing self-defeating (self-undercutting) arguments, which can lead to violations of closure. As an example (free after [Pollock, 1995]) take the argumentation system $(\mathcal{R}_s, \mathcal{R}_d, \mathtt{n}, \leq)$ with $\mathcal{R}_s = \{\rightarrow a;\ b \rightarrow \neg c;\ c \rightarrow \neg b\}$, $\mathcal{R}_d = \{a \Rightarrow b\}$, $\mathtt{n}(a \Rightarrow$

that the properties of non-interference and crash resistance are satisfied. However, the work of Wu and Podlaszewski [2015] assumes that the strength ordering among the defeasible rules is the empty one, and they provide an example of how their approach of erasing inconsistent arguments violates consistency and closure when applying non-empty rule strengths in combination with the last link principle.

The second approach to be discussed is that of Grooters and Prakken [2016]. Here, one of the basic ideas is to change the way strict rules are generated from propositional logic. Instead of generating a strict rule $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ whenever $\varphi_1, \ldots, \varphi_n \vdash \varphi$, they are generating such a strict rule only when from some consistent set $\Phi \subseteq \{\varphi_1, \ldots, \varphi_n\}$ it holds that $\Phi \vdash \varphi$. So instead of the strict rules coinciding with *all* propositional entailment, the idea is to have the strict rules coinciding with *consistent* propositional entailment.

However, ruling out inconsistent inferences alone is not sufficient, as the problem of *ex falso quodlibet* can also occur when successively applying several strict inference steps, as was for instance observed earlier, using the rules $s \rightarrow s \vee r$ and $s \vee r, \neg s \rightarrow \neg r$. The solution proposed by Grooters and Prakken [2016] is simple: when constructing arguments, disallow the application of a strict rule after the application of another strict rule.

It has to be mentioned that the approach of Grooters and Prakken [2016] has not been proven to satisfy any of the properties of direct consistency, closure, indirect consistency, non-interference and crash-resistance. Weaker properties have been proven instead. We refer to [Grooters and Prakken, 2016] for details.

5 Discussion

It is important to observe that the properties examined in the current paper (sometimes called "rationality postulates" in the literature) are not specific to argumentation theory. In fact, they are general properties that can be applied to each formalism for non-monotonic reasoning that aims to encapsulate some form of strict reasoning. This is why the notion of an argument is not mentioned in the postulates of direct consistency, closure, indirect consistency, non-interference and crash-resistance.

$b) = c$ and $\leq = \emptyset$. Here, we can construct arguments $A_1 :\Rightarrow a$, $A_2 : A_1 \Rightarrow b$ and $A_3 : A_2 \rightarrow \neg c$. It holds that A_3 defeats (undercuts) both itself and A_2. This yields a unique complete extension $\{A_1\}$ whose set of conclusions $\{a\}$ satisfies direct consistency, closure and indirect consistency. However, if one were to remove the self-defeating argument A_3, then this would yield a unique complete extension $\{A_1, A_2\}$, whose set of conclusions $\{a, b\}$ violates closure, as it contains b but not $\neg c$. The key point is that whenever one removes a particular class of arguments from the argumentation framework (be it inconsistent or self-attacking arguments) one has to examine whether this results in any violations of direct consistency, indirect consistency and closure.

Instead, these postulates are defined purely based on the *outcome* (in terms of conclusions) of the argumentation formalism. That is, the postulates abstract from the notion of an argument.

This is not to say that no postulates have been formulated specifically about the arguments yielded (instead of about the conclusions yielded). An example of such a postulate would be subargument closure [Caminada and Amgoud, 2007]. This postulate says that if a particular extension contains argument A, then it should also contain all subarguments of A (so each $A' \in Sub(A)$). Satisfying subargument closure is not difficult. From the definition of defeat (under either restricted or unrestricted rebut) it follows that each argument that attacks A' also attacks A. So from A being in, say, a complete extension it follows that A is defended against these attackers, so A' is also being defended. Therefore, A' is also part of the complete extension (which contains everything it defends).

In the current paper, we have mainly focused on rule-based argumentation formalisms, like ASPIC+. However, similar issues also play a role in classical logic based argumentation [Gorogiannis and Hunter, 2011]. Here, the idea is, given a set of propositions Δ (called the *knowledge base*), to construct arguments as pairs $\langle \Phi, \varphi \rangle$ where φ is a proposition (called the *conclusion*) and Φ is a set of propositions (called the *assumptions*) such that $\Phi \vdash \varphi$, $\Phi \nvdash \bot$ and $\neg \exists \phi \in \Phi$: $\Phi \setminus \{\phi\} \vdash \varphi$. Given this argument form, various ways of defining the notion of defeat (or *attack*, as no strength order is taken into account) are examined, especially for their ability to yield a consistent outcome. We refer to the work of Gorogiannis and Hunter [2011] for details. While Gorogiannis and Hunter [2011] do not consider use of preferences, a recent alternative formalisation of classical logic argumentation of D'Agostino and Modgil [2016] satisfies the consistency and non-contamination postulates while supporting the use of preferences. Moreover, this is done without the requirement that an argument's premises need to be checked for consistency and subset minimality, and with the resulting argumentation frameworks only including finite subsets of the arguments defined by a set of classical well-formed formulas. As such, their theory provides a rational account that is suitable for resource bounded agents.

One key point that we want to emphasise is that the satisfaction of rationality postulates is *not* just a matter of theoretical elegance. If we were to apply argumentation theory for practical purposes, to determine what should be the actions to take, and our formalism tells us to put three people on a tandem bicycle, then this advice will be of little use, as the actions to implement it will fail. If we believe the world to be such that there exist some hard (inviolable) constraints, then it makes sense to model these using nondefeasible (strict) rules and expect the argumentation formalism to deal with them in a proper way. Similarly, if one were for instance to build a robot that uses argumentation theory for its internal reasoning,

what we would like to avoid is the situation where after being fed some specific snippets of input (like John whispering in its ear "The cup of coffee contains sugar, and I'm unreliable", and Mary whispering in its ear "The cup of coffee contains no sugar, and I'm unreliable") all inference will come to a grinding halt, and the robot essentially stops functioning. Hence, satisfaction of the rationality postulates is important not just for theoretical elegance, but also to make the theory suitable for actual applications.

Given the important role of rationality postulates when it comes to applications of argumentation theory, we observe that the current state of affairs (at the time of writing) is somewhat unsatisfying. As for the postulates of direct consistency, closure and indirect consistency, there seems to be a dilemma. If, on one hand, one chooses to implement restricted rebut then these postulates can be satisfied under any complete-based semantics. The disadvantage, however, is that restricted rebut can be seen as unintuitive, especially in a dialectical context. If, on the other hand, one chooses to implement unrestricted rebut, then the notion of defeat becomes more in line with natural discussion. The disadvantage, however, is that one can only apply grounded semantics, which tends to yield a very sceptical result. Moreover, satisfaction of the rationality postulates is only guaranteed if the strength order on the defeasible rules is either empty or total (hence ruling out a proper partial oder).

As for the postulates of non-inferference and crash resistance, the situation is even more troublesome. First of all, all the approaches that we are aware of [Wu, 2012; Wu and Podlaszewski, 2015; Podlaszewski, 2015; Grooters and Prakken, 2016] work only with restricted rebut. Moreover, the approach of Wu and Podlaszewski [2015] requires the empty ordering regarding rule strength, whereas in many application domains different rules can have different strengths. The work of Grooters and Prakken [2016], does allow for a non-empty rule strength ordering, but fails to prove any of the forementioned postulates, opting to prove much weaker properties instead.

Overall, when it comes to the development of formal argumentation theory, one can observe that the topic of pure abstract argumentation tends to receive quite some more research attention than the topic of instantiated argumentation. Much work has for instance been done on how to select nodes from a graph. However, the real challenge is how to select nodes from a graph *in a meaningful way*, that is, such that the overall outcome makes sense from a logical perspective so the conclusions could be relied upon regarding what to do or what to believe. If formal argumentation is to be applied in situations that matter, some proper solutions to the issue of rationality postulates would be highly desirable.

Afterword

At the time the current paper was submitted, a paper of Heyninck and Straßer [2017] has just been accepted to be presented at IJCAI 2017. The authors' main idea is to allow for arguments to be attacked on several of its (sub)conclusions (that is, on the conclusions of one or more of its subarguments). This is done by an attacker with a disjunctive conclusion, such that each disjunct is the contrary (negation) of one of the (sub)conclusions of the attacked argument. As far as we know, this yields the first ever instantiation of Dung's argumentation theory that (1) works with a combination of classical logic and defeasible inference rules, (2) satisfies *all* the rationality postulates and (3) implements argument preferences.

References

[Baroni and Giacomin, 2009] P. Baroni and M. Giacomin. Skepticism relations for comparing argumentation semantics. *Int. J. Approx. Reasoning*, 50(6):854–866, 2009.

[Caminada and Amgoud, 2007] M.W.A. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[Caminada and Podlaszewski, 2012] M.W.A. Caminada and M. Podlaszewski. Grounded semantics as persuasion dialogue. In Bart Verheij, Stefan Szeider, and Stefan Woltran, editors, *Computational Models of Argument - Proceedings of COMMA 2012*, pages 478–485, 2012.

[Caminada and Wu, 2009] M.W.A. Caminada and Y. Wu. An argument game of stable semantics. *Logic Journal of IGPL*, 17(1):77–90, 2009.

[Caminada and Wu, 2011] M.W.A. Caminada and Y. Wu. On the limitations of abstract argumentation. In Patrick de Causmaecker, Joris Maervoet, Tommy Messelis, Katja Verbeeck, and Tim Vermeulen, editors, *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011)*, pages 59–66, 2011.

[Caminada *et al.*, 2012] M.W.A. Caminada, W.A. Carnielli, and P.E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254, 2012.

[Caminada *et al.*, 2014a] M.W.A. Caminada, W. Dvořák, and S. Vesic. Preferred semantics as socratic discussion. *Journal of Logic and Computation*, 26:1257–1292, 2014.

[Caminada *et al.*, 2014b] M.W.A. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In Simon Parsons, Nir Oren, Chris Reed, and Drederico Cerutti, editors, *Computational Models of Argument; Proceedings of COMMA 2014*, pages 209–220. IOS Press, 2014.

[Caminada *et al.*, 2015] M.W.A. Caminada, S. Sá, J. Alcântara, and W. Dvořák. On the equivalence between logic programming semantics and argumentation semantics. *International Journal of Approximate Reasoning*, 58:87–111, 2015.

[Caminada, 2014] M.W.A. Caminada. Strong admissibility revisited. In Simon Parsons, Nir Oren, Chris Reed, and Frederico Cerutti, editors, *Computational Models of Argument; Proceedings of COMMA 2014*, pages 197–208. IOS Press, 2014.

[D'Agostino and Modgil, 2016] M. D'Agostino and S. Modgil. A rational account of classical logic argumentation for real-world agents. In *European Conference on Artificial Intelligence (ECAI 2016)*, pages 141–149, 2016.

[Dung and Thang, 2014] P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

[Gorogiannis and Hunter, 2011] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175(9-10):1479–1497, 2011.

[Grooters and Prakken, 2016] D. Grooters and H. Prakken. Two aspects of relevance in structured argumentation: Minimality and paraconsistency. *Journal of Artificial Intelligence Research*, 56:197–245, 2016.

[Heyninck and Straßer, 2017] J. Heyninck and Ch. Straßer. Revisiting unrestricted rebut and preferences in structured argumentation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017. in print.

[Modgil and Prakken, 2013] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intellligence*, 195:361–397, 2013.

[Modgil and Prakken, 2014] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5:31–62, 2014. Special Issue: Tutorials on Structured Argumentation.

[Podlaszewski, 2015] M. Podlaszewski. *Poles Apart: Navigating the Space of Opinions in Argumentation.* PhD thesis, Université du Luxembourg, 2015.

[Pollock, 1995] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person.* MIT Press, Cambridge, MA, 1995.

[Prakken, 2010] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[Toni, 2014] F. Toni. A tutorial on assumption-based argumentation. *Argument & Computation*, 5:89–117, 2014. Special Issue: Tutorials on Structured Argumentation.

[Walton and Krabbe, 1995] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.* SUNY Series in Logic and Language. State University of New York Press, Albany, NY, USA, 1995.

[Wu and Podlaszewski, 2015] Y. Wu and M. Podlaszewski. Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*, 25(2):303–333, 2015.

[Wu, 2012] Y. Wu. *Between Argument and Conclusion; argument-based approaches to discussion, inference and uncertainty.* PhD thesis, Université du Luxembourg, 2012.

 Received 26 May 2017