

On argumentation with purely defeasible rules

Zimi Li¹ and Simon Parsons²

¹ Department of Computer Science, Graduate Center, City University of New York
zli2@gc.cuny.edu

² Department of Computer Science, University of Liverpool
s.d.parsons@liverpool.ac.uk

Abstract. ASPIC⁺ is one of the most widely used systems for structured arguments and includes the use of both strict and defeasible rules. Here we consider using just the defeasible part of ASPIC⁺. We show that using the resulting system, it is possible, in a well defined sense, to capture the same information as using ASPIC⁺ with strict rules.

1 Introduction

Argumentation theory is concerned with the way that intelligent agents discuss whether some statement holds. In the past few years, formal argumentation frameworks have been heavily studied and applications have been proposed in fields such as natural language processing, the semantic web and multi-agent systems. Studying argumentation provides results which help in developing tools and applications in these areas. Dung’s seminal work [8] tells us how to handle the conflicts between arguments. However, it says nothing about the structure of arguments, or how to construct arguments and attack relationships from a knowledge base. Providing the logical basis for argumentation has been the subject of several authors, including [4], [9] and [10]. This paper is concerned with the work that started with the ASPIC [2] framework, and we briefly summarise this work below.

Following [2], [6] pointed out that ASPIC may lead to some non-intuitive results, suggested that all argumentation frameworks must satisfy three rationality postulates in order to avoid these anomalies, and showed how ASPIC could be modified to satisfy them. [13] presented an extension of ASPIC, called ASPIC⁺, which also satisfies the postulates under certain restrictions. [1] and [14] provide further discussion of the approach. [11] modified the ASPIC⁺ framework, to develop a more general structured framework for argumentation with preferences. [5] and [15] presented some examples where ASPIC-like systems could lead to non-intuitive results and gave solutions. Finally, [7] looked at a new variation of ASPIC⁺ which still satisfies the rationality postulates while loosening the restriction on rebut that ASPIC⁺ requires to satisfy the rationality postulates.

Here we continue this line of work, considering another variation of ASPIC⁺ which only contains defeasible elements. We find that, like the system in [7], our system can both loosen the restrictions on ASPIC⁺ and still satisfy the rationality postulates, while being able to establish exactly the same set of conclusions as ASPIC⁺ from the corresponding knowledge-base.

2 Background

2.1 Abstract argumentation

An abstract argumentation framework [8] is a pair $AF = \langle \mathcal{A}, Defeats \rangle$, where \mathcal{A} is a set of arguments, and $Defeats$ is a binary relation collecting all pairs of arguments A and B such that A defeats B , i.e. $Defeats \subseteq \mathcal{A} \times \mathcal{A}$. An argument is called *acceptable* iff it can defend itself, that is, all of its defeaters have been defeated. A subset S of \mathcal{A} is said to be *conflict-free* if there are no arguments in S that defeat an argument in S . Given an abstract argumentation framework, one is typically interested in which of the arguments are acceptable. This is done through argument-based *semantics*, which define different ways to determine acceptability. [8] defines several semantics — complete, grounded, preferred and stable. A given semantics will specify some (possibly empty) sets of acceptable arguments for a given argumentation framework. These sets are also called argument-based extensions, or simply *extensions*. The conclusions of the arguments in an extension are called the *justified conclusions*.

The state-of-the-art way to establish the extensions is through the labeling approach, which is nicely summarized by [3]. This approach can be described in terms of a *labeling function* LF which maps from arguments to a set of labels $\{\text{IN}, \text{OUT}, \text{UNDEC}\}$. Not all labelings are helpful in determining acceptability, and we determine the helpful labelings through the idea of *legality*. For a legal labeling LF , an argumentation framework, $\langle \mathcal{A}, Defeats \rangle$, and an argument $x \in \mathcal{A}$:

1. x is legally IN iff x is labeled IN and every $y \in \mathcal{A}$ that defeats x is labeled OUT.
2. x is legally OUT iff x is labeled OUT and there is at least one $y \in \mathcal{A}$ that defeats x and is labeled IN.
3. x is legally UNDEC iff there is no $y \in \mathcal{A}$ that defeats x such that y is labeled IN, and there is at least one $y \in \mathcal{A}$ that defeats x such that y is labeled UNDEC.

Note that the UNDEC state occurs when x cannot be labeled IN (because it has at least one defeater that is not OUT), and cannot be labeled OUT (because it has no IN defeater). If an argument is not legally labeled, it is said to be *illegally* labeled. More precisely, an argument is illegally labeled l , where $l \in \{\text{IN}, \text{OUT}, \text{UNDEC}\}$, if it is not legally labeled l .

With the notion of legality tying labelings to *Defeats* relations, we can identify acceptable sets of arguments through the notions of *admissibility* and *completeness*. An *admissible* labeling has no arguments that are illegally IN, and no arguments that are illegally OUT. A *complete* labeling is an admissible labeling that, in addition, has no arguments that are illegally UNDEC. Then, given a complete labeling LF , we have: 1) LF is a *grounded* labeling iff there is no complete labeling with a smaller set of IN arguments; 2) LF is a *preferred* labeling iff there is no complete labeling with a larger set of IN arguments; and 3) LF is a *stable* labeling if it contains no UNDEC arguments.

The labeling approach exactly matches Dung's semantics [8]. If LF is a complete labeling, then every x labeled IN by LF is in the complete extension, and so on for grounded, preferred and stable labelings. If an argument is in a given extension, we say that it is justified in the corresponding semantics.

2.2 ASPIC⁺ argumentation framework

Next, we review the ASPIC⁺ argumentation framework in [11]. This defines two kinds of inference rules: strict rules (denoted \rightarrow), meaning the conclusion is always accepted without any exception, and defeasible rules (denoted \Rightarrow), meaning the conclusion is accepted unless there is an exception.

Definition 1 (ASPIC⁺ Argumentation System). *An argumentation system is a triple $AS = \langle \mathcal{L}, \mathcal{R}, n \rangle$ where:*

- \mathcal{L} is a logical language closed under negation $\bar{\cdot}$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict (\mathcal{R}_s) and defeasible (\mathcal{R}_d) inference rules of the form $\phi_1, \dots, \phi_n \rightarrow \phi$ and $\phi_1, \dots, \phi_n \Rightarrow \phi$ respectively (where ϕ_i, ϕ are meta-variables ranging over wff in \mathcal{L}), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \mapsto \mathcal{L}$ is a naming convention for defeasible rules.

We say that a set of propositions in \mathcal{L} is *consistent* iff there do not exist two propositions a and a' such that $a = \bar{a'}$, and it is helpful to think of completing the set of strict rules by considering all the negative connections between propositions mentioned in a strict rule:

Definition 2 (Closure under Transposition). *If \mathcal{R}_s is a set of strict rules, we say \mathcal{R}_s is closed under transposition iff if $\phi_1, \dots, \phi_n \rightarrow \phi \in \mathcal{R}_s$, then $\phi_1, \dots, \phi_{i-1}, \bar{\phi}, \phi_{i+1}, \dots, \phi_n \rightarrow \bar{\phi}_i \in \mathcal{R}_s$ ($i = 1 \dots n$)*

Definition 3 (ASPIC⁺ Knowledge Base). *A knowledge base in an argumentation system $\langle \mathcal{L}, \mathcal{R}, n \rangle$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets \mathcal{K}_n (the axioms) and \mathcal{K}_p (the ordinary premises).*

The above definitions distinguish the premises and the inference rules into two sets, the set of strict elements (\mathcal{R}_s and \mathcal{K}_n) and the set of defeasible elements (\mathcal{R}_d and \mathcal{K}_p).

Definition 4 (ASPIC⁺ Argumentation Theory). *An argumentation theory AT is a pair $\langle AS, \mathcal{K} \rangle$ of an argumentation system AS and a knowledge base \mathcal{K} .*

Before defining precisely what an argument is, we need to introduce some notions which can be defined just understanding that an argument is made up of some subset of the knowledge \mathcal{K} , along with a sequence of rules, that lead to a conclusion. Given this, **Prem**(\cdot) returns all the premises, **Conc**(\cdot) returns the conclusion and **TopRule**(\cdot) returns the last rule in the argument. **Sub**(\cdot) returns all the sub-arguments of a given argument, that is all the arguments that are subset of the given argument.

Definition 5 (Argument). An argument A from of an argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ is:

1. ϕ if $\phi \in \mathcal{K}$ with: $\text{Prem}(A) = \{\phi\}$; $\text{Conc}(A) = \{\phi\}$; $\text{Sub}(A) = \{A\}$; $\text{TopRule}(A) = \text{undefined}$.
2. $A_1, \dots, A_n \rightarrow \phi$ if A_i are arguments such that there exists a strict (or defeasible) rule $\text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow (\text{or } \Rightarrow) \phi$ in \mathcal{R}_s . $\text{Prem}(A) = \text{Prem}(A_1) \cup \dots \cup \text{Prem}(A_n)$; $\text{Conc}(A) = \phi$; $\text{Sub}(A) = \text{Sub}(A_1) \cup \dots \cup \text{Sub}(A_n) \cup \{A\}$; $\text{TopRule}(A) = \text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow (\text{or } \Rightarrow) \phi$.

We write $\mathcal{A}(AT)$ to denote the set of arguments from the theory AT

We say that an argument A is *consistent* iff $\{\text{Conc}(A') \mid A' \in \text{Sub}(A)\}$ is consistent. We further say that an argument A is *strict* if A only contains strict rules, that is $\mathcal{R}_s \neq \emptyset$ and $\mathcal{R}_d = \emptyset$; ³. Similarly, we say: A is *defeasible* if A contains at least one defeasible rule, $\mathcal{R}_d \neq \emptyset$; A is *firm* if A only contains axioms, $\mathcal{K}_n \neq \emptyset$, $\mathcal{K}_p = \emptyset$; A is *plausible* if A contains ordinary premises.

An argument can be attacked in three ways: on its ordinary premises, on its conclusion, or on its inference rules:

Definition 6 (ASPIC⁺ Attack). An argument A attacks an argument B iff A undermines, rebuts or undercuts B , where:

- A undermines B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' = \phi \in \text{Prem}(B)$ and $\phi \in \mathcal{K}_p$.
- A rebuts B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' \in \text{Sub}(B)$ of the form $B'_1, \dots, B'_2 \Rightarrow \phi$.
- A undercuts B (on B') iff $\text{Conc}(A) = \overline{n(r)}$ for some $B' \in \text{Sub}(B)$ such that $\text{TopRule}(B)$ is a defeasible rule r of the form $\phi_1, \dots, \phi_n \Rightarrow \phi$.

We denote “ A attacks B ” by (A, B) .

Note that, in the ASPIC⁺ attack relation, rebutting is *restricted*. That is an argument with a strict **TopRule** can rebut an argument with a defeasible **TopRule**, but not vice versa.

Attacks can be distinguished as to whether they are preference-dependent (rebutting and undermining) or preference-independent (undercutting). The former succeed only when the attacker is preferred. The latter succeed whether or not the attacker is preferred.

Definition 7 (Preference Ordering). A preference ordering \preceq is a binary relation over arguments, i.e., $\preceq \subseteq \mathcal{A} \times \mathcal{A}$, where \mathcal{A} is the set of all arguments constructed from the knowledge base in an argumentation system. We say A ’s preference level is less than or equal to that of B iff $A \preceq B$.

³ this is not same as definition of “strict” as in [11] where the only condition was that $\mathcal{R}_d = \emptyset$. Here we insist that a strict argument includes at least one strict rule. As a consequence, the notions of “strict” and “defeasible” are not duals, and an argument can be neither strict or defeasible — but only if it contains only premises and/or axioms.

In general, neither ASPIC⁺ nor our defeasible system make any assumptions on the properties of the preference ordering, but in establishing a relationship between the two systems, we make use of the *weakest link* principle from [11]. This assumes two pre-orderings \leq, \leq' over \mathcal{R}_d and \mathcal{K}_p respectively, and combines them into $A \prec B$ if:

- the defeasible rules in A include a rule which is weaker than (strictly less than according to \leq) all the defeasible rules in B , and
- the ordinary premises in A include an ordinary premise which is weaker (strictly less than according to \leq') all the ordinary premises in B .

$A \prec B$ is then defined as usual as $A \preceq B$ and $B \not\preceq A$. By combining the definition of arguments, attack relations and preference ordering, we have the following definitions:

Definition 8 (Structured Argumentation Framework). A structured argumentation framework is a triple $\langle \mathcal{A}, att, \preceq \rangle$, where \mathcal{A} is the set of all arguments constructed from the knowledge in the argumentation system, att is the attack relation, \preceq is an preference ordering on \mathcal{A} .

Definition 9 (ASPIC⁺ Defeat). A defeats B iff A undercuts B , or if A rebuts/undermines B on B' and B' 's preference level is less than or equal to that of A ($B' \preceq A$).

Then the idea of an argumentation framework follows from Definitions 5 and 9.

Definition 10 (Argumentation Framework). An (abstract) argumentation framework AF corresponding to a structured argumentation framework $SAF = \langle \mathcal{A}, att, \preceq \rangle$ is a pair $\langle \mathcal{A}, Defeats \rangle$ such that $Defeats$ is the defeat relation on \mathcal{A} determined by SAF .

Example 1 (adapted from [11]). Consider that we have the argumentation system $AS = \langle \mathcal{L}, \mathcal{R}, n \rangle$ where: $\mathcal{L} = \{a, b, c, d, e, f, \bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e}, \bar{f}, \}$, \mathcal{R} is $\mathcal{R}_s = \{d, f \rightarrow \bar{b}\}$ and $\mathcal{R}_d = \{a \Rightarrow b; \bar{c} \Rightarrow d; e \Rightarrow f; a \Rightarrow \bar{nd}\}$ and $n(\bar{c} \Rightarrow d) = nd$. We then add the knowledge-based \mathcal{K} such that $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a; \bar{c}; e; \bar{e}\}$ to get the argumentation theory $AT = \langle AS, \mathcal{K} \rangle$. From this we can construct the arguments:

$$\begin{aligned} A_1 &= [a]; A_2 = [A_1 \Rightarrow b]; A_3 = [A_1 \Rightarrow \bar{nd}]; \\ B_1 &= [\bar{c}]; B_2 = [B_1 \Rightarrow d]; B'_1 = [e]; B'_2 = [B'_1 \Rightarrow f]; B = [B_2, B'_2 \rightarrow \bar{b}]; \\ C &= [\bar{e}]; \end{aligned}$$

Let's call this set of arguments \mathbf{A} , so that: $\mathbf{A} = \{A_1, A_2, A_3, B_1, B_2, B'_1, B'_2, B, C\}$. Note that $\text{Prem}(B) = \{\bar{c}; e\}$, $\text{Sub}(B) = \{B_1; B_2; B'_1; B'_2; B\}$, $\text{Conc}(B) = \bar{b}$, and $\text{TopRule}(B) = d, f \rightarrow \bar{b}$. The attacks between these arguments are shown in Figure 1(a). These make up the set $att = \{(C, B'_1), (B'_1, C), (C, B'_2), (C, B), (B, A_2), (A_3, B_2), (A_3, B)\}$. With a preference order \preceq defined by: $A_2 \prec B; C \prec B; C \prec B'_1; C \prec B'_2$, we have the structured argumentation framework $\langle \mathbf{A}, att, \preceq \rangle$. This structured argumentation framework establishes a defeat relation $Defeats =$

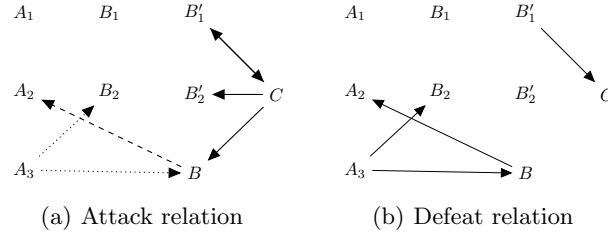


Fig. 1. The attack and defeat relations from Example 1. A dotted arrow shows undercutting, a dashed arrow shows rebutting, and a solid arrow shows undermining.

$\{(B'_1, C), (B, A_2), (A_3, B), (A_3, B_2)\}$ which is shown in Figure 1(b). With this, we can finally write down the argumentation framework $\langle \mathbf{A}, Defeats \rangle$. Note that this is not a rational $ASPIC^+$ framework, since the strict rules are not closed under transposition, but serves to explain the concepts introduced above.

3 $ASPIC_D^+$: a purely defeasible system

3.1 Definition

The full definition of $ASPIC_D^+$ starts from a variation on the $ASPIC^+$ notion of an argumentation system where there are only defeasible elements:

Definition 11 ($ASPIC_D^+$ Argumentation System). An argumentation system is a triple $AS_D = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$ where:

- \mathcal{L} is a logical language closed under negation $\bar{\cdot}$.
- \mathcal{R}_d is a set of defeasible inference rules of the form $\phi_1, \dots, \phi_n \Rightarrow \phi$ (where ϕ_i, ϕ are meta-variables ranging over wff in \mathcal{L}).
- $n : \mathcal{R}_d \mapsto \mathcal{L}$ is a naming convention for defeasible rules.

Definition 12 ($ASPIC_D^+$ Knowledge Base). A knowledge base in an argumentation system $\langle \mathcal{L}, \mathcal{R}_d, n \rangle$ is a set \mathcal{K}_p of ordinary premises.

Definition 13 ($ASPIC_D^+$ Argumentation Theory). An argumentation theory AT_D is a pair $\langle AS, \mathcal{K}_p \rangle$ of an argumentation system AS and a set of ordinary premises \mathcal{K}_p .

Arguments in $ASPIC_D^+$ are then defined as in Definition 5, but there are no strict rules or axioms so there are no strict or firm arguments.

Since any $ASPIC_D^+$ argumentation theory is an $ASPIC^+$ argumentation theory with an empty set of strict rules and an empty set of axioms, we have:

Proposition 1. For a given language \mathcal{L} , \mathbf{AT}_D , the set of all possible $ASPIC_D^+$ argumentation theories, is a subset of \mathbf{AT} , the set of all possible $ASPIC^+$ argumentation theories.

Proof. Pick any ASPIC_D^+ theory $AT_D \in \mathbf{AT}_D$. By definition this is a pair $\langle AS_D, \mathcal{K}_p \rangle$ where $AS_D = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$. It is also an ASPIC^+ theory $AT \in \mathbf{AT}$ where $AT = \langle AS, \mathcal{K}_p \rangle$ (an ASPIC^+ theory with no axioms) and $AS = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$ (an ASPIC^+ theory with no strict rules). Having made no specific assumptions about the composition of AT_D , the result holds for all possible ASPIC_D^+ theories.

However, despite the fact that the set of all possible ASPIC_D^+ theories is a subset of all possible ASPIC^+ theories, we can translate any specific ASPIC^+ theory into a specific ASPIC_D^+ theory. We demonstrate this by defining a translation:

Definition 14 (Defeasible version). ASPIC_D^+ theory AT_D is the defeasible version of ASPIC^+ theory $AT = \langle AS, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ where $AS = \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_p, n \rangle$ iff:

- $AS_D = \langle \mathcal{L}, \mathcal{R}_d \cup \mathcal{R}_{d'}, n' \rangle$, where $\mathcal{R}_{d'} = \{ \phi_1, \dots, \phi_n \Rightarrow \phi \mid \phi_1, \dots, \phi_n \rightarrow \phi \in \mathcal{R}_s \}$ and n' is n extended to name all the rules in $\mathcal{R}_{d'}$.
- $AT_D = \langle AS_D, \mathcal{K}_p \cup \mathcal{K}_{p'} \rangle$, where $\mathcal{K}_{p'} = \{ \phi \mid \phi \in \mathcal{K}_n \}$

If AT_D is the defeasible version of AT , we call AS_D the defeasible version of AS and write $AT_D = \text{def}(AT)$ and $AS_D = \text{def}(AS)$. We call the set of rules $\mathcal{R}_{d'}$ that were strict in AT the set of converted rules, and the set of premises $\mathcal{K}_{p'}$ that were axioms in AT are the set of converted premises. The defeasible version of an argument $A \in \mathcal{A}(AT)$ is an argument $A_D \in \mathcal{A}(AT_D)$ such that every axiom in A is replaced by the corresponding converted premise, and every strict rule in A is replaced by the corresponding converted rule.

In other words, AT_D is the defeasible version of AT , if every axiom of AT becomes an ordinary premise of AT_D , and every strict rule in AT becomes a defeasible rule of AT , while all other components of AT are unchanged.

Given a preference order \preceq over the elements of an ASPIC^+ theory AT , we will need to specify the preference order \preceq_D over the defeasible version of the theory. One way to specify \preceq_D is as follows in terms of the pre-orderings over the rules and premises of AT_D .

Definition 15 (Strict-first preference ordering). Given an ASPIC^+ theory $AT = \langle \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n \rangle, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and preference orders \leq and \leq' over the defeasible rules and premises of that theory, the strict-first preference orderings \leq_{sf} and \leq'_{sf} over the rules and premises of the defeasible version of AT , $AT_D = \langle \langle \mathcal{L}, \mathcal{R}_d \cup \mathcal{R}_{d'}, n' \rangle, \mathcal{K}_p \cup \mathcal{K}_{p'} \rangle$ are such that:

- For every $r, r' \in \mathcal{R}_d$, $r \leq_{sf} r'$ iff $r \leq r'$, and for every $k, k' \in \mathcal{K}_p$, $k \leq'_{sf} k'$ iff $k \leq' k'$.
- For any $r \in \mathcal{R}_d$ and any $r' \in \mathcal{R}_{d'}$, $r <_{sf} r'$, and for every $r', r'' \in \mathcal{R}_{d'}$, $r' =_{sf} r''$.
- For any $k \in \mathcal{K}_p$ and any $k' \in \mathcal{K}_{p'}$, $k <_{sf} k'$, and for every $k', k'' \in \mathcal{K}_{p'}$, $k' ='_{sf} k''$.

where $r =_{sf} r'$ if $r \leq_{sf} r'$ and $r' \leq_{sf} r$, $r <_{sf} r'$ if $r \leq_{sf} r'$ and $r' \not\leq_{sf} r$, $k ='_{sf} k'$ if $k \leq'_{sf} k'$ and $k' \leq'_{sf} k$, and $k <_{sf} k'$ if $k \leq'_{sf} k'$ and $k' \not\leq'_{sf} k$

In other words, all the elements of AT_D that were defeasible in AT have the same preference order as in AT , and all elements that were strict in AT are strictly higher in the preference order than any element that was defeasible in AT . The notion of attack in $ASPIC_D^+$ differs from that in $ASPIC^+$ in that there is no restriction on rebut, and any rule can be undercut:

Definition 16 (ASPIC_D⁺ Attack). *An argument A attacks an argument B iff A undermines, rebuts or undercuts B , where:*

- A undermines B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' = \phi \in \text{Prem}(B)$.
- A rebuts B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' \in \text{Sub}(B)$.
- A undercuts B (on B') iff $\text{Conc}(A) = \overline{n(r)}$ for some $B' \in \text{Sub}(B)$.

With these definitions, we can once again combine the definition of arguments, attack relations and the preference ordering from Definition 7 to get notions of a structured argumentation framework and defeat that are the same as for $ASPIC^+$. To begin to understand the relationship between $ASPIC^+$ and $ASPIC_D^+$, consider this version of Example 1:

Example 2. Consider the $ASPIC_D^+$ argumentation system AS_D which is the defeasible version of the system AS in Example 1. We have $\mathcal{R}_d = \{a \Rightarrow b; \bar{c} \Rightarrow d; e \Rightarrow f; a \Rightarrow \bar{nd}; d, f \Rightarrow \bar{b}\}$, $\mathcal{K}_p = \{a; \bar{c}; e; \bar{e}\}$, and $n(\bar{c} \Rightarrow d) = nd$. We can construct the arguments:

$$\begin{aligned} A_1 &= [a]; A_2 = [A_1 \Rightarrow b]; A_3 = [A_1 \Rightarrow \bar{nd}]; \\ B_1 &= [\bar{c}]; B_2 = [B_1 \Rightarrow d]; B'_1 = [e]; B'_2 = [B'_1 \Rightarrow f]; B = [B_2, B'_2 \Rightarrow \bar{b}]; \\ C &= [\bar{e}]; \end{aligned}$$

Compared with the attacks in Example 1, there is an additional attack here: A_2 rebuts B . With the same preference ordering \preceq over arguments as in Example 1, the defeat relation remains same.

3.2 Properties of $ASPIC_D^+$

We begin by showing that $ASPIC_D^+$ satisfies the three rationality postulates that were introduced in [6] and since then have been considered the basic requirement of a sensible argumentation system. Without strict rules, two of these postulates follow immediately.

Proposition 2 (Closure under Strict Rules). *The conclusions of any extension an $ASPIC_D^+$ theory are closed under strict rules.*

Proof. With no strict rules, the conclusion follows immediately.

Proposition 3 (Direct Consistency). *The conclusions of any extension of an $ASPIC_D^+$ theory are consistent.*

Proof. Suppose the conclusions of one of the extensions E are inconsistent, i.e., there exist two arguments $A, A' \in E$ such that $\text{Conc}(A) = \overline{\text{Conc}(A')}$. If $\text{Conc}(A) \in \mathcal{K}$, by Definition 6, then A' undermines A . On the other hand, if $\text{Conc}(A) \notin \mathcal{K}$, by Definition 6, then A' rebuts A . Either way, A' attacks A in any case. Similarly, A attacks A' . According to Definition 9, at least one of the attack relations is a defeat relation. Therefore, E is not conflict-free and thus E is not an extension under Dung's semantics. The contradiction defeats the assumption of inconsistency and the result holds.

Proposition 4 (Indirect Consistency). *The closure under strict rules of the conclusions of any extension of an ASPIC_D^+ theory is consistent.*

Proof. With no strict rules, this follows immediately from Proposition 3.

Despite the triviality of two of the results, it is worth noting that there are no restrictions on the semantics for which these results hold — they hold for all the standard Dung semantics. Thus ASPIC_D^+ goes further than the ASPIC^- of [7] in extending the scope of reasoning possible with unrestricted rebut since ASPIC^- only satisfied the rationality postulates for the grounded semantics. Of course, this extension is achieved by giving up strict rules, and it is natural to ask what the consequence is for what can be represented in an ASPIC_D^+ theory. Would using ASPIC_D^+ mean any restriction on what can be represented? Our main result is to show that there is no restriction on what can be represented in ASPIC_D^+ compared with what can be represented in ASPIC^+ in the sense that for any ASPIC^+ theory we can build an ASPIC_D^+ theory with the same justified conclusions. We start with the observation that:

Proposition 5. *For a given language \mathcal{L} , there is a defeasible version AT_D of any ASPIC^+ argumentation theory AT .*

Proof. Consider the clauses of Definition 14 as a series of rewrite rules. Any AT can be converted into its defeasible version by turning every axiom into an ordinary premise and every strict rule into a defeasible rule.

This means that whatever information we have in an ASPIC^+ theory, we can capture it in an ASPIC_D^+ theory — we don't lose the ability to represent information about the world by using ASPIC_D^+ rather than ASPIC^+ . However, it is not just representing information that is important. The set of arguments that can be constructed from a theory, and, in particular, the justified conclusions of a theory are also important.

Proposition 6. *Given an ASPIC^+ theory AT and its defeasible version AT_D , $|\mathcal{A}(AT)| = |\mathcal{A}(AT_D)|$ and for every $A \in AT$ there is exactly one $A_D \in \mathcal{A}(AT_D)$ such that A_D is the defeasible version of A .*

Proof. We show there is a 1-to-1 map between $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. For each argument that is just a premise or an axiom $A = [\phi]$, we have $A_D = [\phi]$ that is just a premise; for each argument $A = [A_1, \dots, A_n \Rightarrow \phi]$, we have $A_D = [A_{1D}, \dots, A_{nD} \Rightarrow \phi]$; for each argument $A = [A_1, \dots, A_n \rightarrow \phi]$, we have $A_D = [A_{1D}, \dots, A_{nD} \Rightarrow \phi]$.

Thus any ASPIC^+ theory can be turned into an ASPIC_D^+ theory, and we can generate the same number of arguments, but arguments that had strict components will now only have defeasible components. Furthermore, there are preference orderings such that the same preferences exist between ASPIC_D^+ arguments as between the corresponding ASPIC^+ arguments:

Proposition 7. *Consider the set of arguments \mathcal{A} of an ASPIC^+ theory AT and the set of arguments \mathcal{A}_D constructed from the defeasible version of the theory AT_D . If the preference order over AT_D is the strict-first version of that over AT , then using the weakest link principle, for any $A, B \in \mathcal{A}$, and $A_D, B_D \in \mathcal{A}_D$ where A_D, B_D are the defeasible versions of A and B , $A_D \preceq_D B_D$ iff $A \preceq B$.*

Proof. Let $AT = \langle AS, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and $AT_D = \langle AS_D, \mathcal{K}_{p'} \cup \mathcal{K}_p \rangle$. Consider the preference order \leq over rules in AT , and the preference order \leq' over premises. Let $\langle \leq_D, \leq'_D \rangle$ contain all the relations in $\langle \leq, \leq' \rangle$. Since AF_D has more defeasible elements than AF , we need to determine where these elements fit in the ordering. With a strict-first ordering, the translated strict rules/axioms have the highest preference ordering, and so the weakest links in $\mathcal{A}(AT_D)$ are not the translated strict rules/axioms. Furthermore, all the remaining rules/premises in AT_D have the same preference ordering as in AT . Therefore, AT_D and AT have the same preference ordering over arguments.

In other words, using the weakest link principle, we can take a set of ASPIC^+ arguments create the defeasible versions of those arguments and still have the same preference ordering as over the original set of arguments. This allows us to show our main result, that we can construct a defeasible version of a given ASPIC^+ framework such that the justified conclusions of both theories are the same.

Proposition 8. *Consider a rational ASPIC^+ theory AT and its defeasible version AT_D where the preference ordering over AT_D is the strict-first version of the ordering over AT . Under the weakest link principle, the justified conclusions of AT and AT_D are the same.*

Proof. From Proposition 6 we know that for every argument in $\mathcal{A}(AT_D)$, there is an argument in $\mathcal{A}(AT)$, with the same conclusion, and vice versa. From Proposition 7, we know that under the weakest link principle, the preference order \preceq over $\mathcal{A}(AT_D)$ is the same as the preference order \preceq_D over $\mathcal{A}(AT_D)$. Now, consider the attack relations att and att_D over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. If $(A, A') \in \text{att}$, then $(A_D, A'_D) \in \text{att}_D$ and there is an attack between the defeasible versions of the arguments A_D and A'_D . However, att_D can contains more attacks. (A_D, A'_D) can be in att_D when $(A, A') \notin \text{att}$ iff (1) A' is (just) an axiom in AT or (2) A' has a strict **TopRule** and the attack is not permitted by restricted rebut. We now show, in turn, that these additional attacks do not affect the justified conclusions. First, if A' is an axiom, then A'_D , which as a lone premise that is the defeasible version of an axiom, has the highest possible preference. Thus it can only be defeated by an A_D that has the highest level of preference. Such an argument is

the defeasible version of a strict argument. However, if A was strict, AT would not be rational (it would have two strict elements in conflict). Therefore, we have the same defeat relations over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ and hence the same justified conclusions for AT and AT_D .

Second, if $\text{TopRule}(A')$ is strict, there are two sub-cases that concern us. (a) If $A_D \prec A'_D$, the attack does not become a defeat. Thus AF_D and AF have the same defeat relation, therefore they have the same justified conclusions. (b) If $A'_D \preceq A_D$, then there is one more defeat relation over $\mathcal{A}(AT_D)$ than over $\mathcal{A}(AT)$. We will show that this additional defeat relation has no effect. Consider applying all the defeat relations except this additional one — there are three possibilities for the status of A'_D which will be mirrored by the status of A' which does not have to contend with this additional defeat, and for each of these, we have to consider all three possibilities for the status of A_D .

(1) A'_D is labeled IN. If A_D is labeled IN, then AT has two IN arguments, A and A' , and the conclusions of these arguments are in the set of justified conclusions. However, since A_D and A'_D rebut one another, the conclusions of A and A' are contradictory, violating direct consistency. Thus A and A' cannot both be IN, and so neither can A'_D and A_D before the application of the new defeat. If A_D is labeled OUT then adding the defeat relation (A_D, A'_D) has no effect. If A_D is labelled IN, the situation is more complicated. We start by noting that A will also be UNDEC, and then consider how this can be the case. A has a strict top rule, so $A' = [A'_1, \dots, A'_n \rightarrow a]$ where the top rule is $p_1, \dots, p_n \rightarrow a$. Similarly, $A = [A_1, \dots, A_n \Rightarrow \bar{a}]$ with a top rule $q_1, \dots, q_n \Rightarrow \bar{a}$. By closure under transposition, there exists a strict rule $p_1, \dots, p_{i-1}, \bar{a}, p_{i+1}, \dots, p_n \rightarrow \bar{p}_i$ in AT . Since $A'_D \preceq A_D$, it is not possible for A' to be strict, so A' has at least one defeasible sub-argument, and hence a sub-argument with a defeasible top-rule. Lets assume that this is one of the A'_1, \dots, A'_n that combine with the strict top rule, and call it A'_i . Using the strict rule from the transposition of the top rule, we get an argument $B = [A'_1, \dots, A'_{i-1}, A, A'_{i+1}, \dots, A'_n \rightarrow \bar{p}_i]$ which rebuts A' . B is A plus the A'_j , $j \neq i$, and the transposed strict rule. If A'_1, \dots, A'_n do not have defeasible top rules, then we chain the corresponding transposed strict top rule(s) to B to build an argument that attacks A' further down the argument tree until we get an argument, call it B' , which rebuts A' on its defeasible sub-argument. Now, $A' \preceq B'$ since $A' \preceq A$ and B' is A plus some sub-arguments of A' and a sequence of strict (transposed) rules. Therefore B' defeats A' . Moreover, any defeater of B' must be a defeater of A or A' . Next we consider the labeling. Since A' is labeled IN, all the defeaters of A' are labeled OUT. Since A is labeled UNDEC, all the defeaters of A are labeled OUT or UNDEC. Therefore, the defeaters of B , which are the defeaters of A or A' , are labelled OUT or UNDEC. Thus B is labeled IN or UNDEC. Since B defeats A' , A' can not be labeled IN, contradicting what we started with.

(2) A'_D is labeled OUT. Adding one more defeat relation (A_D, A'_D) has no effect.

(3) A'_D is labeled UNDEC. If A_D is labeled OUT or UNDEC, then adding the defeat relation (A_D, A'_D) has no effect. However, if A_D is labeled IN, then applying the

last defeat relation means that A'_D will now be labeled **OUT** while A' , which does not have to contend with (A, A') , will be **UNDEC**.

So A'_D cannot be initially labeled **IN**. If it is labelled **OUT**, the status of A'_D cannot change as a result of the additional defeat. If A'_D is initially labeled **UNDEC**, the status of A'_D can change. However, by showing that A'_D does not defeat any other arguments we can show that this change does not affect the justified conclusions. Consider an argument $B \in \mathcal{A}(AT_D)$ that is attacked by A'_D . A'_D cannot undercut B since the conclusion of A'_D is not a “rule” (if it were a rule, there would be no rebut between A_D and A'_D and there would be no new defeat relation to consider). A'_D can not undermine B since the conclusion of A'_D is not a premise because we know that A' and hence A'_D has a **TopRule**. So we can only be dealing with a rebut, and since we already know that A_D rebuts A_D , B has to be an argument of which A_D is a sub-argument. Since $A'_D \prec A_D$, A'_D does not defeat B . Thus, in all of these three sub-cases of (b), the additional defeat (A_D, A'_D) has no effect on the status of the arguments in $\mathcal{A}(AT_D)$, again there is no difference between the justified conclusions of AT and AT_D , and the result holds.

This result justifies our claim that ASPIC_D^+ makes it possible to represent the same information as ASPIC^+ . Given an ASPIC^+ theory, we can encode the information in purely defeasible form in an ASPIC_D^+ theory that gives us exactly the same set of (justified) conclusions. The following example helps to show how this is possible.

Example 3. Consider that we start with the following theory AT_1 (given the same language as before), closed under transposition:

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}_s &= \{a' \rightarrow b'; b \rightarrow c; \bar{b}' \rightarrow \bar{a}'; \bar{c} \rightarrow \bar{b}\} \\ \mathcal{K}_n &= \{a; a'\} & \mathcal{K}_p &= \emptyset \end{aligned}$$

Then we can construct the following arguments:

$$\begin{aligned} A_1 &= [a] & A_2 &= [A_1 \Rightarrow b] & A_3 &= [A_2 \rightarrow c] \\ B_1 &= [a'] & B_2 &= [B_1 \rightarrow b'] & B_3 &= [B_2 \Rightarrow \bar{c}] & B_4 &= [B_3 \rightarrow \bar{b}] \end{aligned}$$

The attack relations are shown in Figure 2(a). Now we translate this framework to the ASPIC_D^+ theory AT_2 :

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}'_d &= \{a' \Rightarrow b'; b \Rightarrow c; \bar{b}' \Rightarrow \bar{a}'; \bar{c} \Rightarrow \bar{b}\} \\ \mathcal{K}'_n &= \{a; a'\} & \mathcal{K}_p &= \emptyset \end{aligned}$$

Then we can construct the following arguments:

$$\begin{aligned} A_1 &= [a] & A_2 &= [A_1 \Rightarrow b] & A_3 &= [A_2 \Rightarrow c] \\ B_1 &= [a'] & B_2 &= [B_1 \Rightarrow b'] & B_3 &= [B_2 \Rightarrow \bar{c}] & B_4 &= [B_3 \Rightarrow \bar{b}] \end{aligned}$$

The attack relations are shown in Figure 2(d). Now let's consider the different possible preference orderings over rules:

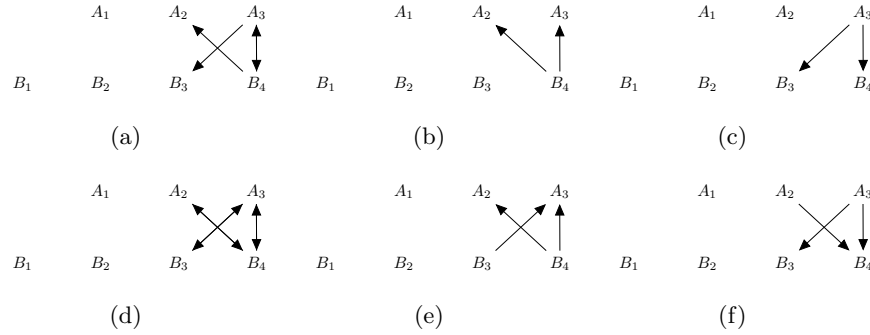


Fig. 2. Attack and defeat relations for AT_1 and AT_2 . (a) Attack relations for AT_1 and Defeat relations for AT_1 in case a), (b) Defeat relations for AT_1 case b), (c) Defeat relations for AT_1 in case c), (d) Attack relations for AT_2 and Defeat relations for AT_2 in case a), (e) Defeat relations for AT_2 case b), (f) Defeat relations for AT_2 case c)

- a) $a \Rightarrow b = b' \Rightarrow \bar{c}^4$. By the weakest link principle, all the attack relations are defeat relations, see Figure 2(a) and 2(d). Here AT_2 has additional defeat relations, but they are directed at arguments that are already defeated. Under the grounded semantics, the set of arguments in the extension is $\{A_1, B_1, B_2\}$ and $\{A_1, B_1, B_2\}$, and the justified conclusions are $\{a, a', b'\}$ and $\{a, a', b'\}$.
- b) $a \Rightarrow b < b' \Rightarrow \bar{c}$. By the weakest link principle, the defeat relations are shown in Figure 2(b) and 2(e). Again AT_2 has an additional defeat relation, but again it has no effect on the justified conclusions. Under the grounded semantics, the set of arguments in the extension is $\{A_1, B_1, B_2, B_3, B_4\}$, $\{A_1, B_1, B_2, B_3, B_4\}$ and $\{A_1, A_2, B_1, B_2, B_3\}$, and the justified conclusions are $\{a, \bar{b}, a', b', \bar{c}\}$ and $\{a, \bar{b}, a', b', \bar{c}\}$.
- c) $a \Rightarrow b > b' \Rightarrow \bar{c}$. By the weakest link principle, the defeat relations are shown in Figure 2(c) and 2(f). As before AT_2 has additional defeats, but they have no effect. Under the grounded semantics, the set of arguments in the extension is $\{A_1, A_2, A_3, B_1, B_2\}$, and $\{A_1, A_2, A_3, B_1, B_2\}$, and the justified conclusions are $\{a, b, a', b', c\}$ and $\{a, b, a', b', c\}$.

For all the cases above, the justified conclusions of AT_1 and AT_2 are exactly same.

4 Conclusion

We have shown that $ASPIC_D^+$, the defeasible subset of $ASPIC^+$, has the same functionality in terms of knowledge representation as $ASPIC^+$. Both formalisms draw the same justified conclusions from corresponding argumentation systems. In addition, $ASPIC_D^+$ goes further than $ASPIC^+$ and $ASPIC^-$ in the sense of satisfying the

⁴ $A = B$ is defined as $A \leq B$ and $B \leq A$

rationality postulates with unrestricted rebut for all of Dung’s semantics. In our view, the work reported here justifies the choice of ASPIC_D^+ in applications such as ArgTrust [12]. Proposition 5 tells us that using ASPIC_D^+ means we can represent exactly the same information that we could in ASPIC^+ , and Proposition 8 tells us that provided that we encode strict rules as defeasible rules with the highest level of preference and use the weakest link principle, the same set of justified conclusion will be obtained as if we had used ASPIC^+ . In addition, we do not have to use restricted rebut, which, as [7] points out, is rather counter-intuitive and hard to explain to users.

References

1. L. Amgoud. Five weaknesses of ASPIC+. In *Advances in Computational Intelligence*, pages 122–131. Springer Verlag, 2012.
2. L. Amgoud, L. Bodestaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. van Veenen, and G. A. W Vreeswijk. Final review and report on formal argumentation system. Deliverable D2.6. Technical report, ASPIC IST-FP6-002307, 2006.
3. P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.
4. P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.
5. M. Caminada. Contamination in formal argumentation systems. In *Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence*, 2005.
6. M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5):286–310, 2007.
7. M. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. *Computational Models of Argument: Proceedings of COMMA 2014*, 2014.
8. P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-persons games. *Artificial Intelligence*, 77(2):321–358, 1995.
9. P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009.
10. A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming*, 4(1+ 2):95–138, 2004.
11. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2012.
12. S. Parsons, E. Sklar, J. Salvit, H. Wall, and Z. Li. Argtrust: decision making with information from sources of varying trustworthiness (Demo). In *Proceedings of the 12th International Conference on Autonomous agents and Multiagent Systems*, 2013.
13. H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.
14. H. Prakken and S. Modgil. Clarifying some misconceptions on the ASPIC+ framework. *Computational Models of Argument: Proceedings of Comma 2012*, 2012.
15. Y. Wu and M. Podlaskowski. Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*, pages 1–31, 2014.