Revisiting Preferences and Argumentation

Sanjay Modgil[1] **and Henry Prakken**[2]

1. Department of Infomatics, King's College London (sanjay.modgil@kcl.ac.uk)

2. Department of Information and Computing Sciences, Utrecht University, and

Faculty of Law, University of Groningen (henry@cs.uu.nl)

Abstract

The *ASPIC*$^+$ framework is intermediate in abstraction between Dung's argumentation framework and concrete instantiating logics. This paper generalises *ASPIC*$^+$ to accommodate classical logic instantiations, and adopts a new proposal for evaluating extensions: attacks are used to define the notion of conflict-free sets, while the defeats obtained by applying preferences to attacks, are exclusively used to determine the acceptability of arguments. Key properties and rationality postulates are then shown to hold for the new framework.

1 Introduction

A Dung *argumentation framework* (*AF*) [Dung, 1995] consists of a binary *attack* relation on a set of arguments. The justified arguments are then evaluated based on the framework's extensions: conflict-free sets of arguments (sets that do not contain arguments that attack) that can be defended against attacks (and so are said to be *acceptable*). The abstract nature of Dung's theory provides for a general and intuitive semantics for the consequence notions of argumentation logics and for nonmonotonic logics in general: an *AF* can be instantiated by the arguments and conflict-based attacks defined by a theory in a logic, and the theory's inferences are then defined in terms of the claims of the justified arguments.

Several works augment *AF*s with preferences and/or values [Amgoud and Cayrol, 2002; Bench-Capon, 2003; Modgil, 2009], so that the conflict-free and acceptable sets of arguments are evaluated only w.r.t the *successful attacks* (*defeats*), where an argument X's attack on Y is successful only if Y is not preferred to X. However, we argue in this paper that it is conceptually more intuitive to continue to define conflict-free sets in terms of those that do not contain attacking arguments. Defeats then encode the preference-dependent success of attacks as they are deployed in the dialectical evaluation of arguments, and so should only be used to determine the acceptability of arguments.

We explore this proposal in the context of the *ASPIC* framework [Amgoud *et al.*, 2006]. The very abstract nature of Dung's framework precludes giving guidance to ensure that the instantiating theory's defined inferences satisfy intuitively rational properties, and so *ASPIC* provided abstract accounts of the structure of arguments, the nature of attack, and the use of preferences. [Caminada and Amgoud, 2007] then formulated consistency and closure postulates that cannot be formulated at Dung's fully abstract level, and showed these postulates to hold for a special case of ASPIC; one in which preferences were *not* accounted for. More recently, *ASPIC*$^+$ [Prakken, 2010] generalised ASPIC and showed that the postulates were satisfied when applying preferences and evaluating the justified arguments on the basis of the derived defeat relation.

This paper makes the following contributions. Firstly, we redefine evaluation of the justified arguments of the *ASPIC*$^+$ framework under our proposed distinct use of attacks and defeats, and show satisfaction of the key properties of Dung frameworks, and [Caminada and Amgoud, 2007]'s rationality postulates. This is significant given that *ASPIC*$^+$ captures a broad range of instantiating logics and argumentation systems, extending those captured by ASPIC (e.g., to include assumption-based argumentation [Bondarenko *et al.*, 1997] and systems using argument schemes). However, *ASPIC*$^+$'s generality is compromised in that it does not capture classical logic approaches to argumentation [Amgoud and Cayrol, 2002; Besnard and Hunter, 2008; Amgoud and Besnard, 2009], since it does not require that the premises of an argument are consistent. This paper's second contribution is to therefore adapt *ASPIC*$^+$ so as to capture these approaches, and so demonstrate satisfaction of the rationality postulates for classical logic approaches that accommodate preferences; a result that to the best of our knowledge has hitherto not been shown for the full range of Dung's original semantics.

The paper is organised as follows. Section 2 reviews background concepts. Section 3 adapts *ASPIC*$^+$ to include classical logic approaches, and defines extensions under the new proposal outlined above. Section 4 shows that key properties of Dung frameworks and the rationality postulates are satisfied [1]. Section 5 discusses related and future work.

2 Background

A *Dung argumentation framework* (*AF*) [Dung, 1995] is a tuple $(\mathcal{A}, \mathcal{C})$, where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation on the arguments $\mathcal{A}$. $S \subseteq \mathcal{A}$ is *conflict free* iff $\forall X, Y \in S$, (X, Y)

[1] Only proofs of the main results are shown in the paper. Proofs not given here can be found in [Modgil and Prakken, 2011].

$\notin \mathcal{C}$. An argument $X \in \mathcal{A}$ is acceptable w.r.t. some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in \mathcal{C}$ implies $\exists Z \in S$ s.t. $(Z, Y) \in \mathcal{C}$. Then:

Definition 1 Let $(\mathcal{A}, \mathcal{C})$ be a *AF*. Then a *conflict free* $S \subseteq \mathcal{A}$ is : an *admissible* extension iff $X \in S$ implies X is acceptable w.r.t. S; a *complete* extension iff $X \in S$ iff X is acceptable w.r.t. S; a preferred extension iff it is a set inclusion maximal complete extension; the grounded extension iff it is the set inclusion minimal complete extension; a stable extension iff it is preferred and $\forall Y \notin S, \exists X \in S$ s.t. $(X, Y) \in \mathcal{C}$.

For $s \in \{$complete, preferred, grounded, stable$\}$, X is *sceptically* or *credulously* justified under the s semantics if X belongs to all, respectively at least one, s extension.

Preference-based *AF*s (*PAF*s) [Amgoud and Cayrol, 2002] are tuples$(\mathcal{A}, \mathcal{C}, \mathcal{P})$, where the preference pre-ordering $\mathcal{P} \subseteq \mathcal{A} \times \mathcal{A}$ determines which attacks succeed as defeats. With the corresponding strict ordering – $Y >_{\mathcal{P}} X$ iff $(Y, X) \in \mathcal{P}$ and $(X, Y) \notin \mathcal{P}$) – then $(X, Y) \in \mathcal{D}$ (i.e., X *defeats* Y) iff $(X, Y) \in \mathcal{C}$ and $Y \not>_{\mathcal{P}} X$. The extensions of $(\mathcal{A}, \mathcal{C}, \mathcal{P})$ are then defined as the extensions of the Dung framework $(\mathcal{A}, \mathcal{D})$.

[Prakken, 2010]'s *ASPIC*$^+$ framework instantiates Dung's abstract approach by assuming an unspecified logical language $\mathcal{L}$, and by defining arguments as inference trees formed by applying strict or defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, interpreted as 'if the *antecedents* $\varphi_1, \ldots, \varphi_n$ hold, then *without exception*, respectively *presumably*, the *consequent* φ holds'. There are two ways to use these rules: they could encode domain-specific information (as in e.g. default logic) but they could also express general laws of reasoning. For example, the defeasible rules could express argument schemes and the strict rules could consist of all classically valid inferences or could more generally conform to any Tarskian consequence notion (cf. [Amgoud and Besnard, 2009]). In order to define attacks, some minimal assumptions on $\mathcal{L}$ are made; namely that certain wff (well formed formulae) are a contrary or contradictory of certain other wff. Apart from this the framework is still abstract: it applies to any set of strict and defeasible inference rules, and to any logical language with a defined contrary relation.

The basic notion of *ASPIC*$^+$ is that of an argumentation system. Arguments are then constructed w.r.t a knowledge base that is assumed to contain three kinds of formulas.

Definition 2 An *argumentation system* is a tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, \leq)$ where:

- $\mathcal{L}$ is a logical language.
- $^-$ is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that:
  - φ is a *contrary* of ψ if $\varphi \in \overline{\psi}$, $\psi \notin \overline{\varphi}$;
  - φ is a *contradictory* of ψ (denoted by '$\varphi = -\psi$'), if $\varphi \in \overline{\psi}$, $\psi \in \overline{\varphi}$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $\leq$ is a preordering on $\mathcal{R}_d$.

Henceforth, a set $S \subseteq \mathcal{L}$ is said to be consistent iff $\nexists\, \psi, \varphi \in S$ such that $\psi \in \overline{\varphi}$, otherwise it is *inconsistent*.

A *knowledge base* in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is a pair $(\mathcal{K}, \leq')$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq'$ is a preordering on $\mathcal{K} \setminus K_n$. Here, $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a$ where these subsets of $\mathcal{K}$ are disjoint: $\mathcal{K}_n$ is the (necessary) *axioms* (which cannot be attacked); $\mathcal{K}_p$ is the *ordinary premises* (on which attacks succeed contingent upon preferences), and; $\mathcal{K}_a$ is the *assumptions* (on which attacks are always successful, c.f. assumptions in [Bondarenko *et al.*, 1997]).

The orderings on defeasible rules and non-axiom premises (we assume their strict counterparts defined in the usual way, i.e., $l < l'$ iff $l \leq l'$ and $l' \not\leq l$) are assumed to be used in defining an ordering $\preceq$ on the constructed arguments (see Section 4). Henceforth, we assume the strict counterpart $\prec$ of $\preceq$ defined in the usual way. Arguments are now defined, where for any argument A, `Prem` returns all the formulas of $\mathcal{K}$ (*premises*) used to build A, `Conc` returns A's conclusion, `Sub` returns all of A's sub-arguments, and `DefRules` returns all defeasible rules in A.

Definition 3 An *argument* A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{Rules}(A) = \emptyset$; $\texttt{TopRule}(A)$ = undefined.
2. $A_1, \ldots A_n \quad \rightarrow\!\!/\Rightarrow \quad \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!/\Rightarrow \psi\}$
   $\texttt{DefRules}(A) = \{r | r \in \texttt{Rules}(A), r \in \mathcal{R}_d\}$
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots \texttt{Conc}(A_n) \rightarrow\!\!/\Rightarrow \psi$

Furthermore, A is: *strict* if $\texttt{DefRules}(A) = \emptyset$; *defeasible* if $\texttt{DefRules}(A) \neq \emptyset$; *firm* if $\texttt{Prem}(A) \subseteq \mathcal{K}_n$; *plausible* if $\texttt{Prem}(A) \not\subseteq \mathcal{K}_n$.

For any $P \subseteq \mathcal{L}$, the closure of P under strict rules, denoted $Cl_{\mathcal{R}_s}(P)$, is the smallest set containing P and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{\mathcal{R}_s}(P)$. Also, we write $S \vdash \varphi$ if there is a strict argument A such that $\texttt{Conc}(A) = \varphi$, with all premises taken from S. Given the intuitive meaning of strict rules, axioms and assumptions, we think that any argumentation system should:

• be *closed under contraposition* or *transposition* (as in [Caminada and Amgoud, 2007]). The former implies that for all $S \subseteq \mathcal{L}$, $s \in S$ and ϕ, if $S \vdash \phi$, then $S \backslash \{s\} \cup \{-\phi\} \vdash -s$. The latter implies that if $\phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{R}_s$, then for $i = 1 \ldots n$, $\phi_1, \phi_{i-1}, -\psi, \phi_{i+1}, \ldots, \phi_n \rightarrow -\phi_i \in \mathcal{R}_s$;

• be *axiom consistent*, i.e., $Cl_{\mathcal{R}_s}(\mathcal{K}_n)$ is consistent.

• be *well formed*, i.e., if φ is a contrary of ψ then $\psi \notin \mathcal{K}_n$ and ψ is not the consequent of a strict rule (since as we will see, attacks by contraries characterise attacks that always succeed; e.g, when ψ is of the from *not* φ in logic programming).

Henceforth, argumentation systems are assumed to satisfy these properties. It should be obvious to see that if the strict rules conform to a Tarskian consequence operator (cf. [Amgoud and Besnard, 2009]), for example, if they consist of all valid propositional or first-order inferences over $\mathcal{L}$, then the first property is always satisfied.

When arguments are inference trees, three syntactic forms of attack from an argument B to an argument A are possible: attacking a premise of A, a conclusion of A, or an inference step in A, respectively called undermining, rebutting and undercutting attacks. To model undercutting attacks on inferences, it is assumed that applications of inference rules can be expressed in the object language; the precise nature of this naming convention will be left implicit. Apart from undercut attacks and attacks on contraries, the success of attacks as defeats is contingent upon preferences.

Definition 4 *A attacks B* iff *A undercuts*, *rebuts* or *undermines B*, where:

• *A undercuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{B'}$ for some $B' \in \texttt{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \psi$.

• *A rebuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' \in \texttt{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \varphi$. In such a case A *contrary-rebuts* B iff $\texttt{Conc}(A)$ is a contrary of φ.

• Argument A *undermines* B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' = \varphi$, $\varphi \in \texttt{Prem}(B) \setminus \mathcal{K}_n$. In such a case A *contrary-undermines* B iff $\texttt{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.

A defeats B iff A undercuts B or A rebuts/undermines B on B' and either A contrary rebuts/undermines B, or $A \not\prec B'$.

3 Argumentation, Logic and Preferences: Revisiting and Generalising *ASPIC*$^+$

As with other works augmenting AFs with *preferences* and/or *values*, *ASPIC*$^+$ evaluates the notions of conflict free and acceptability on the basis of the arguments and *defeats*. [Prakken, 2010] then shows that [Caminada and Amgoud, 2007]'s closure and consistency postulates are satisfied under a number of assumptions. In particular, *direct consistency* is shown by proving that no complete extension yields arguments with conclusions that are the contraries of other arguments. Essentially, this amounts to showing that no complete extension contains arguments that attack. This then begs the question as to why one should not define conflict-free sets as those that do not contain attacking arguments. Intuitively, the success of an attack as a defeat, contingent upon the preference relation, has no bearing on whether an attacking argument is incompatible with the attacked argument, but rather on the dialectical relationship of the former to the latter. Defeats should therefore be reserved for determining whether an attacking argument can be successfully deployed as a counter-argument. That is, they should only be used when determining the acceptability of arguments.

Adopting this new distinct use of attacks and defeats, we now link Section 2's *ASPIC*$^+$ concepts to abstract argumentation frameworks, and define their extensions:

Definition 5 Let $\mathcal{A}$ be a set of arguments as defined in Definition 3. A *structured abstract argumentation framework (SAF)* is a tuple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ such that $(X, Y) \in \mathcal{C}$ iff X attacks Y as defined in Definition 4, and $\preceq$ is a preordering on $\mathcal{A}$.

• Let $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$, where $(X, Y) \in \mathcal{D}$ iff X *defeats* Y as defined in Definition 4

• $S \subseteq \mathcal{A}$ is conflict free iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{C}$.

• The extensions of a *SAF* $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ are the extensions of the Dung framework $(\mathcal{A}, \mathcal{D})$ as defined in Definition 1.

We now also adapt *ASPIC*$^+$ to accommodate classical logic approaches to argumentation [Amgoud and Cayrol, 2002; Besnard and Hunter, 2008]. These assume a propositional or first-order predicate logic and arguments as pairs (S, φ) where S is a consistent and minimal set of wffs that classically entails φ. We therefore identify a class of *SAF* that places restrictions on the instantiating arguments:

Definition 6 A set $S \subseteq \mathcal{L}$ is *c-consistent* if for no φ it holds that $S \vdash \varphi, -\varphi$ (i.e., no strict arguments with contradictory conclusions can be built from premises S). Otherwise S is *c-inconsistent*.

A *SAF* $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ is said to be c-consistent if the arguments $\mathcal{A}$ are as defined in Definition 3, with the added condition that for any $A \in \mathcal{A}$, $\texttt{Prem}(A)$ is c-consistent.

Note that we use the term 'c-consistent' to distinguish the notion of consistency in Definition 2. We can now instantiate c-consistent *SAF*s (c-*SAF*s for short) with arguments defined by classical approaches (we don't need a minimality condition on arguments since Definition 3 guarantees that an argument has no unused premises). The language $\mathcal{L}$ is a first-order language, the contrariness function corresponds to classical negation, and R_d is empty while R_s consists of all valid first-order inferences over $\mathcal{L}$. Furthermore, all elements of a knowledge base are in $\mathcal{K}_p$. Both [Amgoud and Cayrol, 2002; Besnard and Hunter, 2008] consider several notions of attack, one of which corresponds to the present notion of undermining attack.

Henceforth, we assume c-*SAF*s are *c-classical*:

Definition 7 Let $S \subseteq \mathcal{L}$ be a minimal c-inconsistent set iff $\forall S' \subset S$, S' is c-consistent. A c-*SAF* is *c-classical* iff for any minimal c-inconsistent S and any $\varphi \in S$ it holds that $S \setminus \{\varphi\} \vdash -\varphi$ (i.e., amongst all arguments defined there exists a strict argument with conclusion $-\varphi$ with all premises taken from $S \setminus \{\varphi\}$).

If the strict rules in a c-*SAF* conform to a Tarskian consequence operator (cf. [Amgoud and Besnard, 2009]) then it should be obvious to see that the *cSAF* is *c-classical*.

4 Properties of *SAF*s

4.1 Properties of Dung Frameworks

We now prove some fundamental results for *SAF*s and c-*SAF*s. Retaining attacks when defining conflict-free sets potentially undermines some key results shown for Dung frameworks. It may be that A is acceptable w.r.t. an admissible set S, but $S \cup \{A\}$ is not conflict free, so that the *fundamental lemma* [Dung, 1995] does not hold. To illustrate, suppose a *SAF* containing A, B, where B attacks A and $B \prec A$. Then $\{B\}$ is admissible, A is acceptable w.r.t. $\{B\}$, but $\{A, B\}$ is not conflict free. However, under intuitive assumptions on the argument ordering, we show that the result holds. Prior to this we recall some notation from [Prakken, 2010] and then define here the notion of a strict extension of a set of arguments.

Notation 1 Let $M(B)$ denotes the maximal fallible subarguments of B, where for any $B' \in \texttt{Sub}(B)$, $B' \in M(B)$

iff: 1) B' final inference is defeasible or B' is a non-axiom premise, and; 2) there is no $B'' \in \texttt{Sub}(B)$ s.t. $B'' \neq B$ and $B' \in \texttt{Sub}(B'')$ and B'' satisfies 1).

E.g., $M([\Rightarrow r; q; r, q \rightarrow \neg p])$ = $[\Rightarrow r]$ and $[q]$ (assuming $q \in \mathcal{K}_P$ and $\Rightarrow r$ is a defeasible rule with empty antecedent).

Definition 8 Let $\mathcal{A}$ be as defined in Definition 3 or 6. For any $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$, $A \in \mathcal{A}$ is a *strict extension* of $\{A_1, \ldots, A_n\}$ iff:
- the ordinary and assumption premises in A are exactly those in $\{A_1, \ldots, A_n\}$;
- the defeasible rules in A are exactly those in $\{A_1, \ldots, A_n\}$;
- the strict rules and axiom premises of A are a superset of the strict rules and axiom premises in $\{A_1, \ldots, A_n\}$.

Notice that if B defeats some strict extension A of $\{A_1, \ldots, A_n\}$ then the defeat must be on some A_i. Hence:

Lemma 2 Let ($\mathcal{A}$, $\mathcal{C}$, $\preceq$) be a *SAF* or c-*SAF*. Let $A \in \mathcal{A}$ be a strict extension of $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$, and for $i = 1 \ldots n$, A_i is acceptable w.r.t. $E \subseteq \mathcal{A}$. Then A is acceptable w.r.t. E.

We now state under what assumptions a preordering on arguments is said to be *reasonable*.

Definition 9 $\preceq$ is said to be *reasonable* iff:

1. i) $\forall A, B$, if A is strict and firm and B is plausible or defeasible, then $B \prec A$;
   ii) $\forall A, B$, if B is strict and firm then $B \not\prec A$;
   ii) $\forall A, A', B, C$ such that $A \prec B$, $C \prec A$, and A' is a strict extension of A, then $A' \prec B$, $C \prec A'$ (i.e., strict inferences and axiom premises do not change the strength of arguments)
2. Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$, and:
   for $i = 1 \ldots n$, let $C^{+/i}$ be some strict extension of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$.
   Then it is not the case that: $\forall i$, $C^{+/i} \prec C_i$.

The second condition is essentially a weaker form of the property satisfied by any partial ordering defined over a finite set, viz. that there exists a minimal element.

In [Prakken, 2010], an ordering $\prec_s$ is defined over sets: $\Gamma \prec_s \Gamma'$ iff there is a member of Γ that is strictly less than ($<$) all members of Γ', where $<$ is the strict counterpart of the preordering $\leq$ on the defeasible rules or non-axiom premises. Based on $\prec_s$, the commonly used *last* and *weakest* link principles are then used to define example argument orderings $\preceq$. Essentially, $B \prec A$ by the last link principle if the set of top defeasible rules in B is $\prec_s$ the set of top defeasible rules in A, and if both these sets are empty, then the set of non-axiom premises in B is $\prec_s$ the set of non-axiom premises in A. $B \prec A$ by the weakest link principle if the set of defeasible rules in B is $\prec_s$ the set of defeasible rules in A, *and* the set of non-axiom premises in B is $\prec_s$ the set of non-axiom premises in A. We can then show that:

Proposition 3 Let $\preceq$ be defined according to the last or weakest link principle. Then $\preceq$ is *reasonable*.

Observe now that given arguments A = $[\Rightarrow p]$, B_1 = $[\Rightarrow r]$, B_2 = $[\Rightarrow q]$, B = $[B_1; B_2; r, q \rightarrow \neg p]$, then B asymmetrically attacks A, so that if $B \prec A$ then neither argument defeats the other. However, since we assume closure of the strict rules under transposition or contraposition, one can construct strict extensions $A_1^+ = [\Rightarrow p; \Rightarrow q; p, q \rightarrow \neg r]$ of $\{A, B_2\}$ and $A_2^+ = [\Rightarrow p; \Rightarrow r; p, r \rightarrow \neg q]$ of $\{A, B_1\}$, where A_1^+ attacks B on B_1, and A_2^+ attacks B on B_2. Given that $\preceq$ is *reasonable*, and that B is a strict extension of $\{B_1, B_2\}$, then it cannot be that $A_1^+ \prec B_1$, $A_2^+ \prec B_2$ and $B \prec A$. Since by assumption $B \prec A$, then it must be that $A_1^+ \not\prec B_1$ or $A_2^+ \not\prec B_2$, and so either A_1^+ or A_2^+'s attack on B succeeds as a defeat. In fact, the following general result can be shown:

Proposition 4 Let A and B be arguments where B is plausible or defeasible and A and B have contradictory conclusions, and if A and B are defined as in Definition 6, then $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent. Then:

1. For all $B' \in M(B)$, there exists a strict extension $A^+_{B'}$ of $(M(B) \backslash \{B'\}) \cup M(A)$ such that $A^+_{B'}$ rebuts or undermines B on B'.
2. If $B \prec A$, and $\preceq$ is *reasonable*, then for some $B' \in M(B)$, $A^+_{B'}$ defeats B.

We now state some useful results (henceforth $X \rightharpoonup Y$ denotes X *attacks* Y and $X \hookrightarrow Y$ denotes X *defeats* Y).

Lemma 5 Let ($\mathcal{A}$, $\mathcal{C}$, $\preceq$) be a *SAF* or c-*SAF*:

1. If A is acceptable w.r.t. $S \subseteq \mathcal{A}$ then A is acceptable w.r.t. any superset of S.
2. If $A \rightleftharpoons B$ then $A \hookrightarrow B$ or $B \hookrightarrow A$.
3. If $A \hookrightarrow B$, then $A \hookrightarrow B'$ for some $B' \in \texttt{Sub}(B)$, and if $A \hookrightarrow B'$, $B' \in \texttt{Sub}(B)$, then $A \hookrightarrow B$.
4. If A is acceptable w.r.t. $S \subseteq \mathcal{A}$, $A' \in \texttt{Sub}(A)$, then A' is acceptable w.r.t. S.

Lemma 6 Suppose $B \rightharpoonup A$, and not $A \rightharpoonup B$. If not $B \hookrightarrow A$ then either:

1. $\exists A' \in \texttt{Sub}(A)$ s.t. $A' \hookrightarrow B$, or;
2. There is a strict extension $A^+_{B'}$ of $(M(B) \backslash \{B'\}) \cup M(A)$ s.t. $A^+_{B'} \hookrightarrow B$, given that if A and B are defined as in Def. 6, then $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent.

Lemma 7 Let ($\mathcal{A}$, $\mathcal{C}$, $\preceq$) be a c-*SAF*. If $A_1, \ldots, A_n$ are acceptable w.r.t. some conflict-free $E \subseteq \mathcal{A}$, then $\bigcup_{i=1}^{n} \texttt{Prem}(A_i)$ is c-consistent.

Lemma 8 Let A be acceptable w.r.t an admissible extension S of ($\mathcal{A}$, $\mathcal{C}$, $\preceq$). Then $\forall B \in S \cup \{A\}$, neither $B \hookrightarrow A$ or $A \hookrightarrow B$.

We now state the two main results of this section:

Proposition 9 Let A be acceptable w.r.t an admissible extension S of ($\mathcal{A}$, $\mathcal{C}$, $\preceq$). Then S' = $S \cup \{A\}$ is conflict free.

PROOF. Firstly, since for any $B \in S$, B is acceptable w.r.t. S, then by Lemma 7, $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent.

Now, suppose for contradiction that S' is not conflict free.
If $\exists B \in S'$ s.t. $A \rightleftharpoons B$ (this accounts for the case that B = A), then by Lemma 5-2, either $A \hookrightarrow B$ or $B \hookrightarrow A$, contradicting Lemma 8. Else:
1) $\exists B \in S$, $B \rightharpoonup A$, $B \prec A$, and not $A \rightharpoonup B$. By Lemma 6 either:

1.1) some sub-argument A' of A defeats B, hence (by acceptability of B) $\exists C \in S$ s.t. $C \hookrightarrow A'$, and so (by Lemma 5-3) $C \hookrightarrow A$, contradicting Lemma 8, or;
1.2) $\exists\ A^+_{B'}$, $A^+_{B'} \hookrightarrow B$, hence $\exists C \in S$ s.t. $C \hookrightarrow A^+_{B'}$. By construction of $A^+_{B'}$, it must be that $C \hookrightarrow Z$, $Z \in \texttt{Sub}(A) \cup \texttt{Sub}(B)$. Hence, (by Lemma 5-3) either $C \hookrightarrow B$, contradicting S is conflict free, or $C \hookrightarrow A$, contradicting Lemma 8.
2) $\exists B \in S$, $A \rightharpoonup B$, $A \prec B$, and not $B \rightharpoonup A$. By Lemma 6 either:
2.1) some sub-argument B' of B defeats A, hence (by acceptability of A) $\exists C \in S$ s.t. $C \hookrightarrow B'$ and so (by Lemma 5-3) $C \hookrightarrow B$, contradicting S is conflict free, or;
2.2) $\exists B^+_{A'}$ $B^+_{A'} \hookrightarrow A$, hence $\exists C \in S$ s.t. $C \hookrightarrow B^+_{A'}$. By construction of $B^+_{A'}$, it must be that $C \hookrightarrow Z$, $Z \in \texttt{Sub}(A) \cup \texttt{Sub}(B)$, leading to a contradiction as in **1.2)**. QED

Proposition 10 Let A, A' be acceptable w.r.t an admissible extension S of $(\mathcal{A}, \mathcal{C}, \preceq)$. Then:

1. $S' = S \cup \{A\}$ is admissible
2. A' is acceptable w.r.t. S'.

PROOF. 1) By Lemma 5-1, all arguments in S' are acceptable w.r.t. S'. By Proposition 9, S' is conflict free, and hence admissible. 2) By Lemma 5-1, A' is acceptable w.r.t. S'.QED

We have shown that under intuitive assumptions on argument orderings (satisfied by the commonly used weakest and last link principles), and closure of strict rules under transposition or contraposition, *SAF*s and c-*SAF*s satisfy Dung's fundamental lemma (implying, for example, that every admissible extension is a subset of a preferred extension). Also, note that Lemma 5-1 implies monotonicity of the characteristic function, so that a *SAF*'s (c-*SAF*'s) grounded extension can be identified by the function's least fixed point.

4.2 Rationality Postulates

We now show that *SAF*s and c-*SAF*s satisfy [Caminada and Amgoud, 2007]'s rationality postulates for the semantics defined in Definition 1 (it suffices to show they are satisfied by complete extensions).

Theorem 11 *[Sub-argument Closure] Let* $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF *or c-*SAF *and* E *a complete extension of* Δ. *Then for all* $A \in E$*: if* $A' \in \texttt{Sub}(A)$ *then* $A' \in E$.

PROOF. A' is acceptable w.r.t. E by Lemma 5-4, $E \cup \{A'\}$ is conflict free by Prop.9, and so $A' \in E$. QED

Theorem 12 *[Closure under Strict Rules] Let* $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF *or c-*SAF *and* E *a complete extension of* Δ. *Then* $\{\texttt{Conc}(A) | A \in E\} = Cl_{R_s}(\{\texttt{Conc}(A) | A \in E\})$.

PROOF. Follows from Lemma 2 and Prop. 9, where if Δ is a c-*SAF*, Lemma 7 guarantees that A is a valid argument in the sense that its premises are c-consistent. QED

Theorem 13 *[Direct Consistency] Let* $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF *or c-*SAF *and* E *a complete extension of* Δ. *Then* $\{\texttt{Conc}(A) | A \in E\}$ *is consistent.*

PROOF. We show that if $A, B \in E$, $\texttt{Conc}(A) \in \overline{\texttt{Conc}(B)}$, then this leads to a contradiction:

1. A is firm and strict, and:
1.1 if B is firm and strict then this contradicts assumption of *axiom consistency*;
1.2 if B is plausible or defeasible, and **1.2.1** B is an ordinary/assumption premise or has a defeasible top rule, then $A \rightharpoonup B$, contradicting E is conflict free, or **1.2.2** B has a strict top rule (see **3** below).
2. A is plausible or defeasible, and:
2.1 if B is firm and strict then under the *well-formed* assumption $\texttt{Conc}(A)$ cannot be a contrary of $\texttt{Conc}(B)$, and so they are contradictory (i.e., a contrary of each other), and **2.1.1** A is an ordinary/assumption premise or has a defeasible top rule, in which case $B \rightharpoonup A$, contradicting E is conflict free, or **2.1.2** A has a strict top rule (see **3** below);
2.2 if B is plausible or defeasible and **2.2.1** B is an ordinary/assumption premise or has a defeasible top rule then $A \rightharpoonup B$, contradicting E is conflict free, or **2.2.2** B has a strict top rule (see **3** below).
3. Each of **1.2.2**, **2.1.2** and **2.2.2** describes the case where $X, Y \in E$, $\texttt{Conc}(X) \in \overline{\texttt{Conc}(Y)}$, Y is defeasible or plausible and has a strict top rule, and by the *well-formed* assumption $\texttt{Conc}(X)$ and $\texttt{Conc}(Y)$ must be contradictory. In the case that Δ is a c-*SAF*, since $X, Y \in E$, X, Y are acceptable w.r.t. E, and so by Lemma 7, $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent. By Prop 4 there is a strict extension $X^+_{Y'}$ of $M(Y) \backslash \{Y'\} \cup M(X)$ s.t. $X^+_{Y'} \rightharpoonup Y$. By Lemma 2 $X^+_{Y'}$ is acceptable w.r.t. E, and by Prop. 9, $E \cup \{X^+_{Y'}\}$ is conflict free, contradicting $X^+_{Y'} \rightharpoonup Y$. QED

Theorem 14 *[Indirect Consistency] Let* $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF *or c-*SAF *and* E *a complete extension of* Δ. *Then* $Cl_{\mathcal{R}_s}(\{\texttt{Conc}(A) | A \in E\})$ *is consistent.*

PROOF. Follows from Theorems 12 and 13. QED

5 Conclusions

This paper has proposed that in contrast with existing works that obtain defeats by application of preferences, one should retain the attack-based definition of conflict-free sets, given that attacks between arguments indicate the mutual inconsistency of the logical instantiations of the attacking arguments; defeats encode the preference dependent use of attacks in a dialectical context, and so should only be deployed when evaluating the acceptability of arguments. We have reformulated the *ASPIC*$^+$ framework under the new proposal, and generalised *ASPIC*$^+$, thus obtaining structured argumentation frameworks that not only subsume argumentation systems such as [Bondarenko *et al.*, 1997], but also now captures classical logic approaches to argumentation. We have then shown that under some intuitive assumptions, key properties of Dung frameworks, and [Caminada and Amgoud, 2007]'s rationality postulates are satisfied. We are thus the first to demonstrate the consistency of extensions (under *any* semantics subsumed by the complete semantics) obtained by classical logic approaches augmented with preferences [2]. This is

[2]Note that [Amgoud and Vesic, 2010] apply preferences to sets of arguments, rather than to individual attacks, and then show consistency under a variant of stable semantics

of particular importance given that classical-logic approaches without preferences yield (stable/preferred) extensions that simply correspond to the maximal consistent subsets of the instantiating classical logic theories [Cayrol, 1995] (in classical logic preferences thus play a particularly important role in arbitrating between conflicts).

This paper's generalisation and reformulation of *ASPIC*$^+$ treats unsuccessful asymmetric attacks in a different way to [Prakken, 2010]. Such attacks occur when a defeasible or plausible B with strict top rule, rebuts a defeasible A with defeasible top rule, and $B \prec A$. Note that given the intended meaning of strict rules as deductive inferences, it is entirely intuitive that A does not rebut B on the conclusion of B's strict top rule: a basic principle of deductive reasoning is that if one does not like the conclusion of a deductively derived inference, one must *give up* (which in a dialectical setting equates with *deploy an attack against*) one of the defeasible premises. In [Prakken, 2010], the set $\{A, B\}$ is conflict-free, although $\{A, B\}$ cannot be the subset of a complete extension since such an extension would necessarily contain an argument $A^+_{B'}$ that defeats B (as described in Proposition 4). In this paper's version of *ASPIC*$^+$, $\{A, B\}$ is *not* conflict-free, which is conceptually more satisfactory given the logical incompatibility of the information in the arguments.

[Kaci, 2010] argues that to ensure that the use of defeats does not violate consistency, all attack relations should be symmetric, but we agree with [Amgoud and Besnard, 2009] that this would lead to problems. [Amgoud and Vesic, 2009; 2010] claim that unsuccessful asymmetric attacks should result in rejection of the attacker, so that in our example only $\{A\}$ would be admissible. However, our consistency results obviate the need for this. [Amgoud and Vesic, 2010] also motivate their approach by a counterexample to extension consistency but this does not apply to *ASPIC*$^+$ because of its different definition of undermining defeat. Moreover, unlike in *PAF*s [Amgoud and Cayrol, 2002], where all attacks require preferences to be successful, we can make undercutting attack successful irrespective of preferences. This is desirable since unlike with asymmetric rebutting attack, with undercutting attack the attacked argument can, even if R_s is closed, never be extended to yield an attacker of the undercutter, so there is no intuitive sense in which an undercut argument attacks its undercutter. Thus we can, unlike [Amgoud and Vesic, 2009; 2010], preserve the basic principle of argumentation that an un-attacked argument is always justified.

Much work on value based argumentation frameworks (*VAF*s) [Bench-Capon, 2003] considers extensions that contain arguments that asymmetrically attack but do not defeat after accounting for the ranking of the arguments' values. However, little work has been done on logical instantiations of *VAF*s. Since *ASPIC*$^+$ *explicitly* relates the notion of attack to conflict in some (unspecified) instantiating logic, then such extensions are clearly undesirable.

Finally, future work will extend this paper's work in order to structure Extended Argumentation Frameworks (*EAF*s) [Modgil, 2009]. Initial investigations suggest that if one retains the attack-based definition of conflict free for *EAF*s, then the characteristic functions of such structured *EAF*s will be monotonic (in [Modgil, 2009] monotonicity is only shown for the characteristic functions of *hierarchical EAF*s).

References

[Amgoud and Besnard, 2009] L. Amgoud and Ph. Besnard. Bridging the gap between abstract argumentation systems and logic. In *Proc. 3rd International Conference on Scalable Uncertainty (SUM'09)*, pages 12–27, 2009.

[Amgoud and Cayrol, 2002] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[Amgoud and Vesic, 2009] L. Amgoud and S. Vesic. Repairing preference-based argumentation frameworks. In *Proc. of the 21st IJCAI*, pages 665–670, 2009.

[Amgoud and Vesic, 2010] L. Amgoud and G. Vesic. Generalizing stable semantics by preferences. In *Proc. Computational Models of Argument 2010*, pages 39–50. 2010.

[Amgoud *et al.*, 2006] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, J. van Veenen H. Prakken, and G.A.W. Vreeswijk. Final review and report on formal argumentation system. deliverable d2.6, aspic ist-fp6-002307. Technical report, 2006.

[Bench-Capon, 2003] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[Besnard and Hunter, 2008] Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[Bondarenko *et al.*, 1997] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

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

[Cayrol, 1995] C. Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *Proc. of the 14th IJCAI*, pages 1443–1448, 1995.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Kaci, 2010] S. Kaci. Refined preference-based argumentation frameworks. In *Proc. Computational Models of Argument (COMMA) 2010*, pages 299–310. 2010.

[Modgil and Prakken, 2011] S. Modgil and H. Prakken. Revisiting preferences and argumentation: Technical report. http://www.dcs.kcl.ac.uk/staff/smodgil/RPA.pdf, 2011.

[Modgil, 2009] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

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