Resolutions in Structured Argumentation

Sanjay Modgil [a] Henry Prakken [b]
[a] *Department of Informatics, King's College London, UK*
[b] *Department of Information and Computing Sciences, University of Utrecht and Faculty of Law, University of Groningen, The Netherlands*

Abstract. Recently resolution of attacks has been studied in the context of abstract argumentation frameworks. In this paper it is claimed that resolutions should be studied under the assumption that they are generated through the acquisition of preference information, and that this implies that the existing study of resolutions has limited applicability. A formalisation of preference-based resolutions is defined in the context of the *ASPIC*$^+$ framework for structured argumentation, and several properties of resolutions are proven or disproven. It is also argued that when resolutions are modelled without specifying the structure of arguments, then it is easy to overlook that assumptions made at the abstract level do not hold for all reasonable instantiations of the abstract framework, thus reducing its significance.

1. Introduction

The idea of *resolutions* of abstract argumentation frameworks was first introduced in [14]. Given a framework $\Delta = (\mathcal{A}, \mathcal{C})$ (where $\mathcal{A}$ is a set of *arguments* and $\mathcal{C}$ a binary *attack relation* on $\mathcal{A}$), its resolution $\Delta' = (\mathcal{A}, \mathcal{C}')$ is such that $\mathcal{C}'$ replaces one or more symmetric attacks in $\mathcal{C}$ by an asymmetric relation in $\mathcal{C}'$. This was motivated by the application of preferences to symmetric attacks in $\mathcal{C}$, to obtain the *defeat* relation $\mathcal{C}'$. Then properties were suggested that account for the dynamics of preference information; the intuition being that an argument X should be sceptically justified (in all extensions of Δ) iff X is sceptically justified irrespective of how the available preference information is augmented to obtain a resolution Δ' of Δ. This idea was then generalised in [3], where they relate the sceptically justified arguments of frameworks and their resolutions, but without motivating resolutions in terms of extending the existing preference information.

In this paper we argue that in light of [14]'s original motivation, the study of resolutions in [14] and [3] have limited applicability. Specifically, we argue that any such study should account for the use of preferences in obtaining resolutions, and that this is not possible if one restricts resolutions to those defined in [3,14] and if one exclusively models resolutions at the abstract level.

Firstly, we argue that one must also account for the resolution of *asymmetric* attacks, since many argumentation formalisms (e.g., [2,6,15,18]) apply preferences to deny the success of asymmetric attacks as defeats. Furthermore, some formalisms apply preferences so that *both* attacks in a symmetric attack fail to succeed as defeats. We also argue that sometimes resolutions of symmetric attacks are impossible; for example when two symmetrically attacking arguments are assigned equal strength. Resolutions can also be impossible for another reason: preference relations have properties, so the addition of preferences may imply further preferences and thereby make resolutions based on conflicting preferences impossible. Finally, resolutions are impossible if some attacks succeed irrespective of preferences (e.g., attacks on negation as failure assumptions).

In our opinion, such subtleties can only be fully appreciated in a setting where the structure of arguments and the nature of attack and the use of preference to define defeats is made explicit. To this end we study resolutions in the *ASPIC*$^+$ framework [16,17,18], a general framework for argumentation with preferences that integrates and further develops AI models of structured argumentation. In [16,17,18] conditions have been identified under which a range of possible instantiations of *ASPIC*$^+$ satisfy [9]'s rationality postulates, while a number of existing approaches to structured argumentation have been shown to be an instance or special case of *ASPIC*$^+$. Therefore, studying resolutions in this framework arguably makes the study as general as possible.

Section 2 reviews *ASPIC*$^+$, after which Section 3 defines resolutions of *ASPIC*$^+$ argumentation frameworks, defined under the assumption that these are induced by the acquisition of further preference information. Section 4 then evaluates the grounded, preferred and stable semantics against [14]'s properties. We also show that while in general the preferred and stable semantics fail these properties, one can identify specific instantiations of *ASPIC*$^+$ that satisfy them. Sections 3 and 4 also illustrate the above mentioned limitations of considering only resolutions of symmetric attacks, and point to some limits of abstract models of argumentation. In particular, if resolutions are modelled without specifying the structure of arguments, then it is easy to overlook assumptions made at the abstract level that do not hold for all reasonable instantiations of the abstract framework.

2. The *ASPIC*$^+$ framework

We review the *ASPIC*$^+$ framework as defined in [16,17]. *ASPIC*$^+$ assumes an unspecified logical language $\mathcal{L}$, and defines arguments as inference trees formed by applying strict or defeasible inference rules to premises that are well formed formulae (wff) in $\mathcal{L}$. A strict rule means that if one accepts the antecedents, then one must accept the consequent *no matter what*. A defeasible rule means that if one accepts all antecedents, then one must accept the consequent *if there is insufficient reason to reject it*. To define attacks, minimal assumptions on $\mathcal{L}$ are made; namely that certain wff are a contrary or contradictory of certain other wff, and that defeasible inference rules can be named in the language $\mathcal{L}$ through the use of a naming convention n. 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 an argumentation system. Arguments are then constructed w.r.t a knowledge base that is assumed to contain three kinds of formulas.

Definition 1 An ***ASPIC*$^+$ argumentation system** is a tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ where $\mathcal{L}$ is a logical language and $^-$ 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 of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), and such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.

$n : \mathcal{R}_d \longrightarrow \mathcal{L}$ is a naming convention for defeasible rules.

$\leq$ is a partial pre-ordering on $\mathcal{R}_d$.

Definition 2 An **ASPIC**$^+$***knowledge base*** in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ is a pair $(\mathcal{K}, \leq')$ where:

- $\mathcal{K} \subseteq \mathcal{L}$, and $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a$ where these subsets of $\mathcal{K}$ are disjoint, and: $\mathcal{K}_n$ is the (necessary) *axioms* ; $\mathcal{K}_p$ is the *ordinary premises*; $\mathcal{K}_a$ is the *assumptions*.
- $\leq'$ is a partial pre-order on the non-axiom premises $\mathcal{K} \setminus K_n$.

Intuitively, axiomatic premises cannot be attacked, the success of attacks (as defeats) on ordinary premises is contingent on preferences, while attacks on assumptions always result in defeats (*cf.* assumptions in [7]).

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 Rules returns all rules in A.

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

1) φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{Rules}(A) = \emptyset$.

2) $A_1, \ldots A_n \rightarrow / \Rightarrow \psi$ 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 $\mathcal{R}_s / \mathcal{R}_d$.
$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$; $\texttt{Conc}(A) = \psi$;
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;
$\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow / \Rightarrow \psi\}$,
$\texttt{DefRules}(A) = \{r | r \in \texttt{Rules}(A), r \in \mathcal{R}_d\}$

A is *strict* if it contains no defeasible rules (i.e., $\texttt{DefRules}(A) = \{r | r \in \texttt{Rules}(A), r \in \mathcal{R}_d\} = \emptyset$); *defeasible* otherwise. Also, A is firm if $\texttt{Prem}(A) \subseteq \mathcal{K}_n$; plausible otherwise.

In [16], the notion of *c-consistent* arguments is also introduced, so that classical logic approaches to argumentation (e.g., [2,13]) can be captured as instances of *ASPIC*$^+$. These approaches require that the premises of arguments are consistent. Hence, a c-consistent argument A is an argument whose premises cannot be extended by strict rules to obtain arguments with contradictory conclusions (i.e., one cannot on the basis of $\texttt{Prem}(A)$ construct strict arguments B and B' s.t. $\texttt{Conc}(B) = \varphi$, $\texttt{Conc}(B') = -\varphi$).

Three kinds of *attack* are defined for *ASPIC*$^+$ arguments. B can attack A by attacking a premise or conclusion of A, or an inference step in A. Some kinds of attack are *preference-independent* in that they result in *defeats* independently of preferences over arguments. Other kinds of attack are instead *preference-dependent* in that they result in *defeats* only if the attacked argument is not stronger than the attacking argument (see [17] for a detailed discussion of the rationale for this distinction). The orderings on defeasible rules and non-axiom premises (we assume their usual strict counterparts, i.e., $l < l'$ iff $l \leq l'$ and $l' \nleq l$) may be used in defining an ordering $\preceq$ on the constructed arguments (we also assume the strict counterpart $\prec$ of $\preceq$). For example, in [17,18] $\preceq$ is defined according to the weakest or last link principles[1].

[1]Informally, the last-link principle compares arguments in terms of their last-applied defeasible rules or (if there are no such rules) in terms of their non-axiom premises, while the weakest link principle compares arguments both in terms of all their defeasible rules and their non-axiom premises.

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

• A *undercuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.

• 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 $\varphi \in \mathcal{K}_a$.

An undercut, contrary-rebut, or contrary-undermine attack is said to be *preference-independent*, otherwise an attack is *preference-dependent*.

Then, A **defeats** B (denoted $A \rightarrow B$) iff A attacks B (denoted $A \rightharpoonup B$) on B', and either: $A \rightharpoonup B'$ is *preference-independent*, or; $A \rightharpoonup B'$ is *preference-dependent* and $A \nprec B'$.

The success of rebutting and undermining attacks thus involves comparing the conflicting arguments at the points where they conflict. The definition of successful undermining exploits the fact that an argument premise is also a subargument.

Adding an argument ordering (which may or may not be defined on the basis of the partial preorders on defeasible rules and non-axiom premises) to an argumentation theory consisting of an argumentation system and a knowledge base, yields a structured argumentation framework:

Definition 5 Let AT be an **argumentation theory** (AS, KB). A **structured argumentation framework (*SAF*)** defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all, or only c-consistent, arguments constructed from KB in AS, $\preceq$ is a partial preorder on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

The justified arguments under Dung semantics [11] can then be defined. To recap, a Dung framework consists of a binary relation $\mathcal{B}$ over a set of arguments $\mathcal{A}$. Then:

- $S \subseteq \mathcal{A}$ is *conflict free* iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{B}$;
- $X \in \mathcal{A}$ is acceptable w.r.t. some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in \mathcal{B}$ implies $\exists Z \in S$ s.t. $(Z, Y) \in \mathcal{B}$.

Then, a conflict free set S is:

- an *admissible* extension iff $X \in S$ implies X is acceptable w.r.t. S;
- a *complete* extension iff $X \in S$ whenever 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. $(X, Y) \in \mathcal{B}$.

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.

If $\Delta = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ is a *SAF*, and $\mathcal{D}$ the defeat relation obtained from $\mathcal{C}$ and the preference ordering $\preceq$, then letting $\mathcal{D}$ be the binary relation $\mathcal{B}$, the justified arguments of Δ are the justified arguments of the Dung framework $(\mathcal{A}, \mathcal{D})$. In [18] it is shown that under some intuitive assumptions on the strict inference rules, axiom premises and the preference relation $\preceq$, instantiations of *ASPIC*$^+$ satisfy [9]'s rationality postulates for argumentation.

In [16,17], it is argued that unlike [18] and other works that derive defeats from attack relations, conflict free-ness should be defined w.r.t. attacks ($\mathcal{B} = \mathcal{C}$ in the above definition of conflict-free), and the defeat relation should only be used to define the acceptability of arguments ($\mathcal{B} = \mathcal{D}$ in the above definition of acceptability). [16,17] then show that under the above mentioned assumptions, the key results for Dung's theory are preserved, and [9]'s rationality postulates are satisfied. Henceforth we will assume evaluation of justified arguments as defined in [16,17].

In summary, [16,17,18] show that *ASPIC*$^+$ provides a general structured account of argumentation with preferences. We identify conditions under which a range of possible instantiations satisfy [9]'s rationality postulates, and formally show that the framework captures a number of existing approaches to argumentation, including assumption based argumentation [7], Carneades [12] (as proven in [20]), and instances of Tarskian [1] (and in particular classical [2,13]) logic approaches extended with preferences.

3. Preference-based Resolutions in *ASPIC*$^+$

Abstract argumentation frameworks are instantiated by logical formalisms where an attack from X to Y denotes that the claim of X is in a relationship of logical conflict with an element in the support (or conclusion) of Y, thus underpinning the definition of conflict free. An attack from X to Y additionally licenses its use in the dialectical evaluation of the acceptability of an argument. The former denotation of an attack is fully determined by the instantiating logic and contrariness relation (that generalises negation), so that questioning its validity amounts to questioning the logical axiomatisation of conflict [2] (demarcating these distinct roles of attacks motivates [16]'s retention of attacks when defining conflict free sets in *ASPIC*$^+$). It follows then, that the notion of a resolution should be interpreted under the understanding that the removal of an attack equates with denying its dialectical use. The latter has been extensively studied in the context of argumentation with preferences (e.g., [2,6,15]), where X's attack on Y fails to succeed as a defeat if Y is preferred to X. We therefore argue that the notion of a framework and its resolutions should be studied under the assumption that the latter are generated by augmenting the preference information in the former, and, as will be shown, any such study must account for the structure of arguments and the nature of attacks. As discussed in Section 1, this implies that the study of semantics with respect to the extensions generated by purely *abstract* frameworks and resolutions in which only symmetric attacks are resolved [3,14] has limited applicability. Note that a similar criticism also applies to the resolution based semantics of [5].

We are therefore interested in the case where given a SAF $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ and its defined defeat relation, what is the relationship, under different semantics, between the sceptically justified arguments of Δ and the sceptically justified arguments of $\Delta' = (\mathcal{A}, \mathcal{C}, \preceq')$, where Δ' is a resolution of Δ obtained by extending' $\preceq$ to the preference relation $\preceq'$.

[2]More informal instantiations may sanction disagreement regarding relationships of conflict. However, this paper assumes the established understanding of abstract argumentation theory as a formalism for instantiation by logical formalisms, as originally conceived in [11].

Definition 6 Let $\preceq$ be a partial preorder over a set Γ. Then $\preceq'$ *extends* $\preceq$ iff $\preceq \subseteq \preceq'$ and $\forall X, Y \in \Gamma$, $X \prec Y$ implies $X \prec' Y$.
Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a *SAF*. Then $\Delta' = (\mathcal{A}, \mathcal{C}, \preceq')$ *preference-extends* Δ iff $\preceq'$ *extends* $\preceq$.

To motivate the definition of *extends*, recall that $\preceq$ is a partial preorder. Thus it does not suffice to define *extends* in terms of the condition $X \prec Y$ implies $X \prec' Y$ alone. To see why, suppose $X \preceq Y$ and $Y \preceq X$, which implies $X \approx Y$; that is they are effectively assigned the same strength. Hence, it might be that $\preceq'$ preserves the strict preferences in $\preceq$, but $X \npreceq Y$ and $Y \npreceq X$. But we certainly want to preserve the assignment of equal strength to X and Y. On the other hand, it does not suffice to define *extends* in terms of the condition $\preceq \subseteq \preceq'$ alone. This is because given only $X \preceq Y$ and so $X \prec Y$, we want that this strict preference be preserved in the extended argument ordering. However, if $X \preceq' Y$ and $Y \preceq' X$, then this strict preference would not be preserved.

It is straightforward to then show that if $(\mathcal{A}, \mathcal{C}, \preceq')$ *preference-extends* $(\mathcal{A}, \mathcal{C}, \preceq)$, and $\mathcal{D}'$ and $\mathcal{D}$ are the defeat relations respectively defined by $\preceq'$ and $\preceq$, then $\mathcal{D}' \subseteq \mathcal{D}$. Furthermore:[3]

Proposition 1 Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be defined by $(\mathcal{K}, \leq_1)$ in $(\mathcal{L}, ^-, \mathcal{R}, \leq_2)$, and $\preceq$ defined on the basis of $\leq_1$ and $\leq_2$ according to the weakest or last principles (as defined in Section 5.1 in [17]). For any $\leq'_1$ that extends $\leq_1$, and any $\leq'_2$ that extends $\leq_2$, the SAF Δ' = $(\mathcal{A}, \mathcal{C}, \preceq')$ defined by $(\mathcal{K}, \leq'_1)$ in $(\mathcal{L}, ^-, \mathcal{R}, \leq'_2)$, *preference-extends* Δ.

We can now define the notion of a preference-based resolution:

Definition 7 Let $\Delta' = (\mathcal{A}, \mathcal{C}, \preceq')$ be a SAF that *preference-extends* $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$, and let $\mathcal{D}'$ and $\mathcal{D}$ be defeat relations respectively defined by $\preceq'$ and $\preceq$. Then Δ' is a *preference-based resolution* of Δ iff $\mathcal{D}' \subset \mathcal{D}$.

The resolutions defined here differ from the resolutions in [3,14], which only resolve symmetric relations between arguments. Firstly, preferences may result in denying the dialectical success of an asymmetric attack as a defeat, both in abstract approaches to argumentation (e.g., Value Based Frameworks [6]), but also in structured approaches. For example, in *ASPIC*$^+$ an argument with a strict top rule can asymmetrically attack an argument with a defeasible top rule while being weaker than its target, so that the attack does not result in a defeat. Also, classical logic approaches to argumentation with preferences [2] (and their formulation in *ASPIC*$^+$ [16,17]) define only asymmetric attacks from the conclusion of an attacking argument to the premise of the attacked argument. Secondly, some resolutions may not be possible. If X and Y in Δ are equally strong ($X \approx Y$), and X and Y attack and so defeat each other, then any further preferences, and thus any further resolution, preserves the assignment of equal strength and thus the symmetric defeat. For example, if two contradicting witnesses or experts were deemed to be equally credible, then no further preferences can change this. Likewise, if two conflicting laws were regarded of equal hierarchical status then no further preferences can change this. Thirdly both attacks in a symmetric attack may fail to succeed as defeats, as illustrated by the following example [4].

[3] Space limitations preclude inclusion of proofs for all but the key results in this paper. All results not formally proven in this paper, can be found in Section 9, [17]

[4] Results of [16,17,18] imply that [9]'s consistency postulates hold for this and all our further examples.

Example 2 Let $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ be an argumentation system where:

- $\mathcal{L}$ is a language of propositional literals, composed from a set of propositional atoms $\{a, b, c, \dots\}$ and the symbols $\neg$ and $\sim$ respectively denoting strong and weak negation (i.e., negation as failure). α is a strong literal if α is a propositional atom or of the form $\neg\beta$ where β is a propositional atom. α is a wff of $\mathcal{L}$, if α is a strong literal or of the form $\sim \beta$ where β is a strong literal.
- For any wff α, α and $\neg\alpha$ are contradictories and α is a contrary of $\sim \alpha$.
- $\mathcal{R}_s = \emptyset$, $\mathcal{R}_d = \{\neg c \Rightarrow \neg b; a, b \Rightarrow c\}$, and $\leq$ = $\approx$ (since partial pre-orders are reflexive $\leq$ = $\approx$ denotes $\{r \leq r | r \in \mathcal{R}_d\}$)

$(\mathcal{K}, \leq')$ is the knowledge base such that $\mathcal{K}_n = \emptyset$, $\mathcal{K}_p = \{a, b, \neg c\}$, $\mathcal{K}_a = \emptyset$, and $\leq$ = $\{a <' \neg c <' b\}$.

We obtain arguments X = $[\neg c; \neg c \Rightarrow \neg b]$ and Y = $[a; b; a, b \Rightarrow c]$. Then X attacks Y on Y' = $[b]$, and Y attacks X on X' = $[\neg c]$. Under the `Elitist` ordering on sets whereby $S \prec_{\texttt{Eli}} S'$ if an element in S is ordered below all elements in S', we have the following orderings on non-axiom premises: $\{\neg c\} \prec_{\texttt{Eli}} \{b\}$ and $\{a, b\} \prec_{\texttt{Eli}} \{\neg c\}$. These orderings respectively determine (by the weakest or last link principles as defined in [17]) that $X \prec Y'$ and $Y \prec X'$. Hence neither X or Y defeat each other.

4. Evaluating Semantics Against Properties of Preference-based Resolutions

[14] states properties which capture the intuition that a sceptical reasoner should consider an argument to be justified iff it is justified irrespective of how her current preference information may be extended. We restate these properties for preference-based resolutions, and then evaluate the grounded, preferred and stable semantics with respect to them [5].

Property 3 [**Left to Right Sceptical**] If X is a sceptically justified argument of Δ = $(\mathcal{A}, \mathcal{C}, \preceq)$, then X is a sceptically justified argument of every preference-based resolution Δ' = $(\mathcal{A}, \mathcal{C}, \preceq')$ of Δ.

Property 4 [**Right to Left Sceptical**] If X is a sceptically justified argument of every preference-based resolution Δ' = $(\mathcal{A}, \mathcal{C}, \preceq')$ of Δ = $(\mathcal{A}, \mathcal{C}, \preceq)$, then X is a sceptically justified argument of Δ.

The grounded semantics fails *Right to Left Sceptical*, as illustrated by the following:

Example 5 Let $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ be the argumentation system where $\mathcal{L}$ and the contrary relations are defined as in Example 2, and:

- $\mathcal{R}_s = \emptyset$, $\mathcal{R}_d = \{\neg q \Rightarrow p; \neg p \Rightarrow q; \sim p, \sim q \Rightarrow r; \sim r \Rightarrow s\}$, and $\neg q \Rightarrow p \leq \neg p \Rightarrow q$ and $\neg p \Rightarrow q \leq \neg q \Rightarrow p$ (i.e., $\neg q \Rightarrow p \approx \neg p \Rightarrow q$)
- $(\mathcal{K}, \leq')$ is the knowledge base $\mathcal{K}_n = \emptyset$, $\mathcal{K}_p = \{\neg p, \neg q\}$, $\mathcal{K}_a = \{\sim p, \sim q\}$, and $\leq'$ = $\approx$

Figure 1-a) shows the induced arguments and defeats ($\preceq$ = $\approx$ given that $\leq'$ = $\approx$). Note the attacks on R and S are contrary attacks and so are preference independent, and since α is a contrary of $\sim \alpha$, the arguments $[\sim p]$ and $[\sim q]$ do not attack and so defeat P and Q

[5] Properties 3 and 4 are generalised respectively to *skepticism adequacy* and *resolution adequacy* in [3].

respectively. Figures 1-b) and 1-c) show the two possible preference-based resolutions, obtained respectively by extending $\leq'$ to include $\neg q <' \neg p$ (and so $P \prec P'$, $P \prec Q$, $Q' \prec Q$) and $\neg p <' \neg q$ (and so $P' \prec P$, $Q \prec P$, $Q \prec Q'$). Argument S is in the grounded extension of both resolutions, but not in the grounded extension of Figure 1-a).

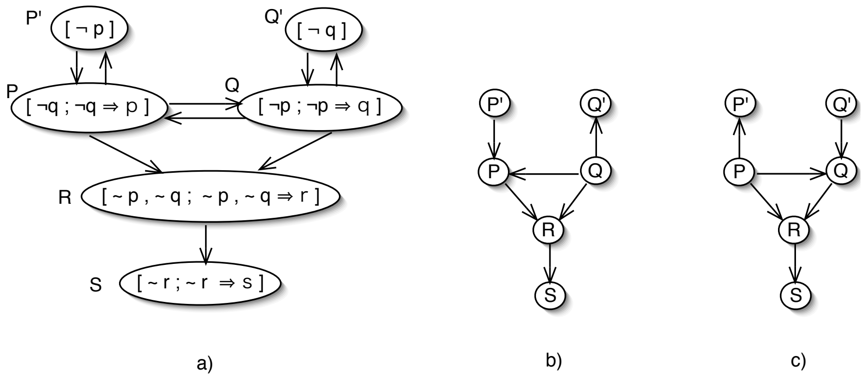

Figure 1. b) and c) are the two preference-based resolutions of a)

However, one can prove *Left to Right Sceptical* for finitary frameworks:

Theorem 6 *If X is in the grounded extension of $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$, then X is in the grounded extension of every preference-based resolution Δ' of Δ.*

PROOF. *Let $\mathcal{D}$ and $\mathcal{D}'$ be the defeat relation defined by Δ and Δ' respectively.*
Let $\bigcup_{i=1}^{n} F_i$ be the grounded extension obtained by iterative application of the characteristic function $\mathcal{F}$ to $(\mathcal{A}, \mathcal{D})$ (i.e., $F_1 = \mathcal{F}(\emptyset)$, $F_i = \mathcal{F}(F_{i-1})$).
Let $\bigcup_{i=1}^{m} G_i$ be the grounded extension obtained by iterative application of the characteristic function $\mathcal{F}$ to $(\mathcal{A}, \mathcal{D}')$.
We show by induction on i, that $X \in F_i$ implies $X \in G_i$:
Base case (i =1) *: $F_1 = \{X | \neg\exists Y, (Y, X) \in \mathcal{D}\}$. Hence, since $\mathcal{D}' \subset \mathcal{D}$, $G_1 \subseteq F_1$.*
Inductive Hypothesis *: For $j < i$, $X \in F_j$ implies $X \in G_j$.*
General Case *: Suppose $X \in F_i$, $Y \rightarrow_{\mathcal{D}} X$. Then $\exists Z \in F_{i-1}$, $Z \rightarrow_{\mathcal{D}} Y$. By inductive hypothesis $Z \in G_{i-1}$. Suppose $Y \rightarrow_{\mathcal{D}'} X$, $Z \nrightarrow_{\mathcal{D}'} Y$ (since $Z \prec' Y$). By Lemma 34 in [17], either $\exists Y' \in \mathtt{Sub}(Y)$ s.t. $Y' \rightarrow_{\mathcal{D}'} Z$, or $\exists Y^{+}_{Z'}$ s.t. $Y^{+}_{Z'} \rightarrow_{\mathcal{D}'} Z$ on some defeasible sub-argument Z' of Z ($Y^{+}_{Z'}$ is an argument strictly extending the defeasible sub-arguments of Y and all the defeasible arguments of Z except Z'). Since $Z \in G_{i-1}$, then $\exists W \in G_j, j < i$, $W \rightarrow_{\mathcal{D}'} Y$.*

QED

The preferred and stable semantics fail *Left to Right Sceptical*. To illustrate, consider the argumentation system where $\mathcal{L}$ and the contrary relations are defined as in Example 2, and the arguments (built from assumption premises) and defeats are shown in Figure 2-a). We assume no ordering on the defeasible rules and non-axiom premises, and so $\preceq$ = $\approx$ (recall that attacks on assumption premises are preference independent). $\{D, B\}$ is the single preferred/stable extension and so set of sceptically justified arguments. Now, consider the preference-based resolution in Figure 2-b) obtained by adding the ordering

$\sim e \Rightarrow \neg a < \sim c \Rightarrow a$, and so $D \prec A$ (under the last link principle as defined in Section 5.1 [17]). The single preferred/stable extension of this resolution is $\emptyset$.

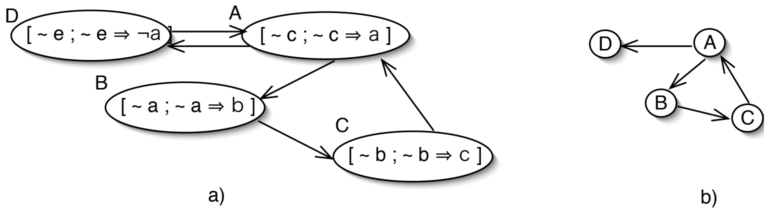

Figure 2. b) is a preference-based resolution of a)

In [14] and subsequently [3], it is shown that *Right to Left Sceptical* holds for the preferred semantics, for resolutions as defined in [3,14]. We now show that once we account for preference-based resolutions, *Right to Left Sceptical* fails, even in the case where such resolutions only resolve symmetric attacks.

Example 7 Consider the argumentation system where $\mathcal{L}$ and the contrary relations are defined as in Example 2, and where:
• $\mathcal{R}_s = \emptyset$, $\mathcal{R}_d = \{\sim b, \sim c \Rightarrow a, \sim a \Rightarrow x\}$.
• The knowledge base consists of $\mathcal{K}_n = \emptyset$, $\mathcal{K}_p = \{\neg b, b, \neg c, c\}$, $\mathcal{K}_a = \{\sim b, \sim c, \sim a\}$, and $\neg b <' c$, $\neg c <' b$.
Based on either the weakest or last link principles, $D \prec C$, $E \prec B$. We obtain the arguments and defeats shown in Figure 3-a). $\{D, E, A\}$ is one of the preferred/stable extensions, and so X is not sceptically justified. We now enumerate all possible ways of extending the ordering $\leq'$ on the ordinary premises, and thus (by Proposition 1) $\preceq$ (defined under either the weakest or last link principles), and the resultant resolutions. Note that extending the ordering on defeasible rules will make no difference as only the attacks between B and D, and E and C are preference dependent):

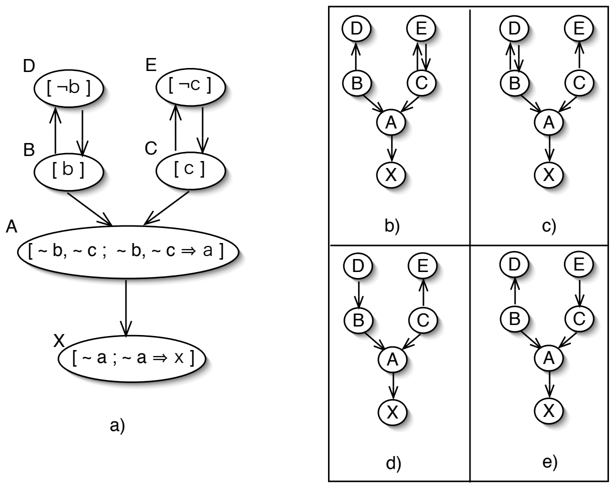

Figure 3. b) is a preference-based resolution of a)

1. Extending with $\neg b <' b$ yields $D \prec B$ and the resolution in Figure 3-b). The preferred/stable extensions are $\{B, E, X\}$ and $\{B, C, X\}$.
2. Extending with $\neg c <' c$ yields $E \prec C$ and the resolution in Figure 3-c). The preferred/stable extensions are $\{C, D, X\}$ and $\{B, C, X\}$.
3. Extending with $\neg b <' b$ and $\neg c <' c$ yields $D \prec B$, $E \prec C$, and the resolution (not shown) with preferred extension $\{B, C, X\}$
4. Extending with $b <' \neg b$, then by transitivity $\neg c <' c$, yielding $B \prec D$, $E \prec C$, and the resolution in Figure 3-d). The preferred/stable extension is $\{C, D, X\}$.
5. Extending with $c <' \neg c$, then by transitivity $\neg b <' b$, yielding $D \prec B$ and $C \prec E$, and the resolution in Figure 3-e). The preferred/stable extension is $\{B, E, X\}$
6. Extending with $\neg b <' \neg c$, then by transitivity $\neg b <' b$, and we are in case 1.
7. Extending with $\neg c <' \neg b$, then by transitivity, $\neg c <' c$, and we are in case 2.
8. Extending with $b <' c$, then by transitivity $\neg c <' c$, and we are in case 2.
9. Extending with $c <' b$, then by transitivity $\neg b <' b$, and we are in case 1.

The counter-example thus shows that for all preference-based resolutions, X is a sceptically justified argument. However X is not a sceptically justified argument of Δ.

Example 7 illustrates the impossibility of constructing a resolution Δ' with D asymmetrically defeating B and E asymmetrically defeating C, which would yield a preferred/stable extension $\{D, E, A\}$ that excludes X, and thus would preserve *Right to Left Sceptical*. The only reason *Right to Left Sceptical* holds for preferred semantics in [3,14] is that in the abstract setup all resolutions are possible, including Δ'. However the *ASPIC*$^+$ instantiation (cases 4 and 5) illustrates that given the existing premise ordering, any extension making $B \prec D$ (and so $D \rightarrow B$) then implies $E \prec C$ (and so $C \rightarrow E$), and any extension making $C \prec E$ ($E \rightarrow C$) then implies $D \prec B$ ($B \rightarrow D$).

The results for preferred/stable semantics are negative. However, the question naturally arises as to whether particular *ASPIC*$^+$ instantiations satisfy the desired properties, and under what restrictions. In what follows we show that the properties are satisfied by particular classical logic instantiations of *ASPIC*$^+$.

Consider an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ where $\mathcal{L}$ is a standard propositional or first-order language, $^-$ is defined as classical negation, $\mathcal{R}$ consists only of strict inference rules $\mathcal{R}_s$ which consists of all valid first-order inferences over $\mathcal{L}$, and $\leq$ = $\approx$. Let $(\mathcal{K}, \leq')$ be any knowledge base with $\mathcal{K}_n = \mathcal{K}_a = \emptyset$, $\mathcal{K}_p = \Gamma$, $\Gamma \subseteq \mathcal{L}$, and $\leq'$ a total preorder over $\mathcal{K}_p$ (Γ). Let Δ = $(\mathcal{A}, \mathcal{C}, \preceq)$ where $\mathcal{A}$ is the set of c-consistent arguments and $\preceq$ defined under the weakest or last link link principle. We write $\Delta_{(\Gamma, \leq')}$ to denote such a *SAF*. Then, for the stable semantics, *Left to Right Sceptical* and *Right to Left Sceptical* can be shown by exploiting an equivalence (Theorem 32 in [17]) between the above classical logic instantiation of *ASPIC*$^+$ and Brewka's *preferred subtheories* [8]:

Definition 8 A default theory is a tuple $(\Gamma, \leq)$, where Γ is a set of classical first order formulae, $\leq$ is a total pre-order and $(\Gamma_1, \ldots, \Gamma_n)$ the $\leq$ induced partition into equivalence classes, such that $\forall \alpha, \beta \in \Gamma$, $\alpha < \beta$ iff $\alpha \in \Gamma_i, \beta \in \Gamma_j$, $i > j$.
A preferred subtheory is a set Σ = $\Sigma_1 \cup \ldots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \ldots \cup \Sigma_i$ is a maximal (under set inclusion) consistent subset of $\Gamma_1, \ldots, \Gamma_i$. Henceforth we write $\mathcal{PS}((\Gamma, \leq))$ to denote the set $\{\Sigma_1, \ldots, \Sigma_n\}$ of all preferred subtheories of Γ.

Intuitively, a preferred subtheory is obtained by taking a maximal under set inclusion consistent subset of Γ_1, extending this with a maximal consistent subset of Γ_2, extending this with a maximal consistent subset of Γ_3, and so on.

Theorem 8 *Let* $\Delta_{(\Gamma,\leq)}$ = $(\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF*, and for any* $\Sigma \subseteq \Gamma$*, let* $\texttt{Args}(\Sigma) \subseteq \mathcal{A}$ *be the set of all arguments with premises taken from* Σ*. Then:*
1) If Σ *is a preferred subtheory of* $(\Gamma, \leq)$*, then* $\texttt{Args}(\Sigma)$ *is a stable extension of* $\Delta_{(\Gamma,\leq)}$*.*
2) If E *is a stable extension of* $\Delta_{(\Gamma,\leq)}$*, then* $\bigcup_{A \in E} \texttt{Prem}(A)$ *is a preferred subtheory of* $(\Gamma, \leq)$*.*

We now present the following properties of preferred subtheories (formally proven in Section 9, [17]) in which we say that a default theory $(\Gamma, \leq')$ extends $(\Gamma, \leq)$ iff $\leq'$ extends $\leq$ (as defined in Definition 6).

Proposition 9 Let $(\Gamma, \leq')$ be any default theory extending $(\Gamma, \leq)$. Then Σ is a preferred subtheory of Γ' implies Σ is a preferred subtheory of Γ.

Proposition 10 Let $(\Gamma, \leq'_1), \ldots, (\Gamma, \leq'_n)$ be all the default theories extending $(\Gamma, \leq)$. Suppose that for $i = 1 \ldots n$, $\alpha \in \cap\mathcal{PS}((\Gamma'_i, \leq'_i))$. Then $\alpha \in \cap\mathcal{PS}((\Gamma, \leq))$.

Theorem 11 *Let* $\Delta_{(\Gamma,\leq)}$ = $(\mathcal{A}, \mathcal{C}, \preceq)$ *be a* SAF*. Then* Left to Right Sceptical *and* Right to Left Sceptical *are satisfied under the stable semantics.*

PROOF. Left to Right Sceptical*: Suppose* $X \in \mathcal{A}$ *sceptically justified. By Theorem 8-2), for every preferred subtheory* Σ *of* $(\Gamma, \leq)$*,* $\texttt{Prem}(X) \subseteq \Sigma$*. Assume some extension* $(\Gamma, \leq')$*,* Σ' *a preferred subtheory of* $(\Gamma, \leq')$*,* $\texttt{Prem}(X) \nsubseteq \Sigma'$*. Then by Proposition 9,* Σ' *is a preferred subtheory of* $(\Gamma, \leq)$*. By Theorem 8-1),* $\texttt{Args}(\Sigma')$ *is a stable extension, and by assumption* $X \notin \texttt{Args}(\Sigma')$*, contradicting* X *is sceptically justified.*
Right to Left Sceptical*: Suppose* X *sceptically justified in any preference-based resolution* $\Delta'_{(\Gamma,\leq')}$ = $(\mathcal{A}, \mathcal{C}, \preceq')$*. By Theorem 8-2),* $\texttt{Prem}(A) \subseteq \cap\mathcal{PS}((\Gamma, \leq'))$*. By Proposition 10,* $\texttt{Prem}(X) \subseteq \cap\mathcal{PS}((\Gamma, \leq))$*. By Theorem 8,* $\Sigma \in \mathcal{PS}((\Gamma, \leq))$ *iff* $Args(\Sigma)$ *is a stable extension of* Δ*. Hence* X *is in every stable extension of* Δ*.* QED

5. Conclusions

We have argued that resolutions and the related properties should more properly be studied under the assumption that resolutions are induced by extensions of the preference relation, and that any such study must account for the structure of arguments. Our work points to the limited applicability of existing approaches to resolutions which do not allow for removal of asymmetric relations, and do not allow for the fact that some preference-based resolutions may not be possible.

Properties proven at the abstract level [3,14] are disproved here, since the abstract approaches make assumptions that do not hold for instantiations (in particular that all resolutions are always possible). That our instantiations do not satisfy these assumptions is not, we argue, to be viewed as an anomaly, given that all the example argumentation theories described in this paper satisfy [9]'s rationality postulates, and our modelling of preference-based resolutions seems entirely plausible. This paper echoes [19]'s similar critiques of preference-based abstract argumentation frameworks [2], and so further supports the *ASPIC*$^+$ view that a general model of preference handling in argumentation cannot be given at the abstract level, but must make the structure of arguments explicit. Furthermore, the limitations we show for abstract accounts of resolutions, apply more generally to any abstract account of adding or deleting attacks or arguments (e.g., [4,10]).

All this work implicitly assumes that such additions or deletions are independent of each other, an assumption that may not hold for instantiations.

We conclude by pointing to future work. The instance of *ASPIC*$^+$ that is shown in Section 4 to satisfy both properties, suggests investigating other conditions under which properties are satisfied. Also, the intuitions underlying the properties, suggest other properties by which semantics could be evaluated. For example, 'X is a credulously justified argument of Δ iff X is a sceptically justified argument of some preference-based resolution Δ'', and the weaker postulate 'X is a credulously justified argument of Δ iff X is a credulously justified argument of some preference-based resolution Δ''.

References

[1] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *Proc. 4th International Conference on Scalable Uncertainty (SUM'10)*, pages 42–55, 2010.

[2] 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.

[3] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(1015):675 – 700, 2007.

[4] R. Bauman and G. Brewka. Expanding Argumentation Frameworks: Enforcing and Monotonicity Results. In *Proc. 2010 Conference on Computational Models of Argument (COMMA 2010)*, 75–86, 2010.

[5] P. Baroni, P. E. Dunne, and M. Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3-4):791–813, 2011.

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

[7] 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.

[8] G. Brewka. Preferred subtheories: an extended logical framework for default reasoning. In *Proc. 11th International Joint Conference on Artificial intelligence*, pages 1043–1048, 1989.

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

[10] C. Cayrol, F. Dupin de Saint-Cyr and M-C. Lagasquie-Schiex Revision of an Argumentation System. In *Proc. Knowledge Representation (KR'08)*, 124-134, 2008.

[11] 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.

[12] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

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

[14] S. Modgil. Hierarchical argumentation. In *Proc. 10th European Conference on Logics in Artificial Intelligence (JELIA)*, pages 319–332, 2006.

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

[16] S. Modgil and H. Prakken. Revisiting preferences and argumentation. In *IJCAI 2011*, 2011.

[17] S. Modgil and H. Prakken. A general account of argumentation and preferences. Technical report, www.dcs.kcl.ac.uk/staff/smodgil/GAP.pdf, 2012.

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

[19] H. Prakken. Some reflections on two current trends in formal argumentation. In A. Artikis, et.al (eds.) *Logic Programs, Norms and Action. Essays in Honour of Marek J. Sergot on the Occasion of his 60th Birthday*, pp. 249–272. Springer, 2012.

[20] B. van Gijzel and H. Prakken. Relating Carneades with abstract argumentation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 1113–1119, 2011.