Reasoning about Preferences in Structured Extended Argumentation Frameworks

Sanjay Modgil[a], Henry Prakken[b] [1]
[a]*Department of Computer Science, University of Liverpool,* [b]*Department of Information and Computing Sciences, Utrecht University & Faculty of Law, University of Groningen*

Abstract. This paper combines two recent extensions of Dung's abstract argumentation frameworks in order to define an abstract formalism for reasoning about preferences in structured argumentation frameworks. First, extended argumentation frameworks extend Dung frameworks with attacks on attacks, thus providing an abstract dialectical semantics that accommodates argumentation-based reasoning *about* preferences over arguments. Second, a recent extension of the ASPIC framework (ASPIC+) instantiates Dung frameworks with accounts of the structure of arguments, the nature of attack and the use of preferences to resolve attacks. In this paper, ASPIC+ is further developed in order to define attacks on attacks, resulting in a dialectical semantics that accommodates argumentation based reasoning about preferences in structured argumentation. Then, some recently proposed rationality postulates for structured extended argumentation are proven to hold.

Keywords. Abstract Argumentation, Preferences, Postulates.

1. Introduction

A Dung *argumentation framework* (DF) [6] consists of a binary *attack* relation on a set of arguments. The justified arguments are then evaluated under different semantics. The abstract nature of *DFs* successfully provides for a general and intuitive semantics for the consequence notions of argumentation logics and for nonmonotonic logics in general: a DF can be instantiated by the arguments and 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. On the other hand, the abstract nature of DFs precludes giving guidance as to what kinds of instantiation ensure that the instantiating theory's defined inferences satisfy intuitively rational properties. To address this issue, the ASIPC abstract framework for structured argumentation [2] integrated work on rule-based argumentation [12,15,14] with [6]'s abstract approach. ASPIC provides abstract accounts of the structure of arguments, the nature of attack, and the use of a given preference ordering to determine which attacks succeed as *defeats*. [5] then exploited this added expressiveness to formulate several consistency and closure rationality postulates that cannot be formulated at Dung's fully abstract level. These postulates were then proven to be satisfied for a special case of [2]'s ASPIC framework; one in which preference orderings were *not* accounted for.

[1]Corresponding Author: Sanjay Modgil, E-mail: sanjaymodgil@yahoo.co.uk.

More recently, [13] generalised the ASPIC framework to develop ASPIC+. The significance of this work is that: 1) ASPIC+ is proven to capture a broader range of systems than ASPIC, e.g., assumption-based argumentation [4] and systems using argument schemes; 2) ASPIC+, and so any existing or new argumentation logic instantiating ASPIC+, is shown to satisfy [5]'s postulates for the more general case in which preferences *are* accounted for. Hence, for example, preferences can be applied to remove attacks defined by an instantiation of ASPIC+, whilst guaranteeing that the claims of the arguments in a complete extension are mutually consistent.

In a parallel development, [7] addressed a limitation of developments of *DF*s that account for the relative strengths of attacking arguments in order to determine which attacks succeed as defeats [1,3]. While [1] and [3] respectively assume *given* preference and value orderings for valuating the relative strengths of arguments, in reality, such valuations are often themselves the outcome of argumentation based reasoning. To model this, [7] extends *DF*s so that arguments expressing preferences attack the attacks between the arguments over which the preferences are expressed. [7] then defines evaluation of the justified arguments of these Extended Argumentation Frameworks (*EAFs*) under each of the Dung semantics. However, while [7] investigates two specific instantiations of *EAFs*, a general principled account of structured argumentation accommodating argumentation about preferences has thus far been lacking.

In this paper we provide such an account. Section 2 reviews Dung's theory, *EAFs*, and ASPIC+. Section 3 then builds on ASPIC+ to allow for arguments that express preferences over other arguments, and which then instantiate a version of [7]'s *EAFs* in the same general way as *DFs* have been instantiated by ASPIC+. We then show that the resulting *structured EAF*s satisfy [5]'s rationality postulates. The significance of this work is that it enables principled development of novel and existing instantiating logics (e.g., [4]) to incorporate reasoning about priorities; principled in the sense that these logics' inferences, defined now through instantiation of *structured EAF*s, are guaranteed to satisfy [5]'s rationality postulates. For example, one can now guarantee that despite the fact that the instantiating logic defines arguments that attack and so remove attacks, the claims of arguments contained in a complete extension of the instantiated *EAF* will be mutually consistent. In Section 3.3 we illustrate this with an example instantiation.

2. Background

2.1. A Review of Abstract Argumentation

A *Dung argumentation framework* (*DF*) [6] is a tuple $(\mathcal{A}, \mathcal{C})$, where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation on the arguments in $\mathcal{A}$. An argument $X \in \mathcal{A}$ is then said to be 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. $(Z, Y) \in \mathcal{C}$ (i.e., Z *reinstates* X). A *DF*'s characteristic function $\mathcal{F}$ is defined such that for any $S \subseteq \mathcal{A}$, $\mathcal{F}(S)$ = $\{X | X$ is acceptable w.r.t. S $\}$. We now recall Dung's definition of extensions under different semantics:

Definition 1 Let $(\mathcal{A}, \mathcal{C})$ be a *DF*, $S \subseteq \mathcal{A}$ be *conflict free* (i.e., $\forall X, Y \in S$, $(X, Y) \notin \mathcal{C}$): S is an *admissible* extension iff $S \subseteq \mathcal{F}(S)$; S is a *complete* extension iff $S = \mathcal{F}(S)$; S is a preferred extension iff it is a set inclusion maximal complete extension; S is a grounded extension iff it is a set inclusion minimal complete extension (since $\mathcal{F}$ is monotonic there

is guaranteed to be a unique grounded extension given by $\mathcal{F}$'s least fixed point); S is 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$\}$, $X \in \mathcal{A}$ is *sceptically* justified under the s semantics, if X belongs to all s extensions, and *credulously* justified if X belongs to at least one s extension.

Extended Argumentation Frameworks (*EAF*s) [7] extend *DF*s to include a second attack (*pref-attack*) relation:

Definition 2 [*EAF*] An *EAF* is a tuple ($\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$), where ($\mathcal{A}, \mathcal{C}$) is a *DF*, $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{C}$, and if $(Z, (X, Y))$, $(Z', (Y, X)) \in \mathcal{D}$ then (Z, Z'), $(Z', Z) \in \mathcal{C}$.

Note the constraint on any Z, Z', where given that they respectively pref-attack (X, Y) and (Y, X), then they express contradictory preferences (Y is preferred to X, respectively X is preferred to Y) and so themselves symmetrically attack.

Henceforth, we focus on *bounded hierarchical EAF*s that are stratified so that attacks at some level i are only pref-attacked by arguments in the next level up (such *EAF*s have been shown to suffice for many applications of *EAF*s [8,9,10]):

Definition 3 [*bh-EAF*s] $\Delta = (\mathcal{A}, \mathcal{C}, \mathcal{D})$ is a *bounded hierarchical EAF* (*bh-EAF*) iff there exists a partition $\Delta_H = (((\mathcal{A}_1, \mathcal{C}_1), \mathcal{D}_1), \ldots, ((\mathcal{A}_n, \mathcal{C}_n), \mathcal{D}_n))$ such that $D_n = \emptyset$, and:

- $\mathcal{A} = \bigcup_{i=1}^{n} \mathcal{A}_i$, $\mathcal{C} = \bigcup_{i=1}^{n} \mathcal{C}_i$, $\mathcal{D} = \bigcup_{i=1}^{n} \mathcal{D}_i$, and for $i = 1 \ldots n$, $(\mathcal{A}_i, \mathcal{C}_i)$ is a *DF*.
- $(C, (A, B)) \in \mathcal{D}_i$ implies $(A, B) \in \mathcal{C}_i$, $C \in \mathcal{A}_{i+1}$.

The notion of a successful attack (*defeat*) is then parameterised w.r.t. preferences specified by some given set S of arguments:

Y *defeats*$_S$ X, denoted $Y \rightarrow^S X$, iff $(Y, X) \in \mathcal{C}$ and $\neg\exists Z \in S$ s.t. $(Z, (Y, X)) \in \mathcal{D}$.

An *EAF conflict free* set S is then defined as a set that does not admit arguments that symmetrically attack, but can contain some Y and X such that Y *asymmetrically* attacks X, given a $Z \in S$ that pref-attacks this attack. That is, S is *conflict free* iff :

$\forall X, Y \in S$: if $(Y, X) \in \mathcal{C}$ then $(X, Y) \notin \mathcal{C}$, and $\exists Z \in S$ s.t. $(Z, (Y, X)) \in \mathcal{D}$.

The acceptability of an argument X w.r.t. a set S requires that there is a *reinstatement set* for any reinstating defeat:

Definition 4 [EAF acceptability] Let $S \subseteq \mathcal{A}$ in ($\mathcal{A}$, $\mathcal{C}$, D). Let $R_S = \{X_1 \rightarrow^S Y_1, \ldots, X_n \rightarrow^S Y_n\}$ where for $i = 1 \ldots n$, $X_i \in S$. Then R_S is a reinstatement set for $A \rightarrow^S B$, iff $A \rightarrow^S B \in R_S$, and $\forall X \rightarrow^S Y \in R_S$, $\forall Y'$ s.t. $(Y', (X, Y)) \in \mathcal{D}$, $\exists X' \rightarrow^S Y' \in R_S$

X is acceptable w.r.t. $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $Y \rightarrow^S X$, there is a *reinstatement set* for some $Z \rightarrow^S Y$.

Given this definition of acceptability, admissible, preferred, complete, grounded and stable extensions are defined in the same way as for DFs (except that '$X \rightarrow^S Y$' replaces '$(X, Y) \in \mathcal{C}$' in the definition of stable extensions), and Dung's fundamental lemma [6] is shown to hold for *EAF*s. The domain of an *EAF*s characteristic function $\mathcal{F}$ is limited to conflict free sets and is monotonic for *bh-EAF*s, so that the grounded extension is defined by the least fixed point of $\mathcal{F}$ [2].

[2] For arbitrary *EAF*s, $\mathcal{F}$ is not monotonic. However [7] shows that iterating $\mathcal{F}$ starting from the empty set does provide a fixed point that identifies the grounded extensions

2.2. A Framework for Structured Argumentation

As stated earlier, the ASPIC+ framework of [13] further develops [2,5]'s instantiation of [6]'s abstract frameworks with accounts of the structure of arguments, the nature of attack and the use of preferences to resolve attacks. The framework instantiates Dung's abstract approach by assuming an unspecified logical language and by defining arguments as inference trees formed by applying strict or defeasible inference rules. The notion of an argument as an inference tree naturally leads to three ways of attacking an argument: attacking an inference, attacking a conclusion and attacking a premise. To resolve such conflicts, preferences may be used, which leads to three corresponding kinds of defeat: undercutting, rebutting and undermining defeat. To characterise them, some minimal assumptions on the logical object language are made; namely that certain well-formed formulas are a contrary or contradictory of certain other well-formed formulas. Apart from this the framework is still abstract: it applies to any set of inference rules divided into strict and defeasible, and to any logical language with a defined contrary relation.

The basic notion of [13]'s framework is that of an argumentation system.

Definition 5 [Argumentation system] 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 if $\varphi \in \overline{\psi}$ then if $\psi \notin \overline{\varphi}$ then φ is called a *contrary* of ψ, otherwise φ and ψ are called *contradictory*. The latter case is denoted by $\varphi = -\psi$ (i.e., $\varphi \in \overline{\psi}$ and $\psi \in \overline{\varphi}$).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ 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 partial preorder 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*.

Arguments are built by applying inference rules to one or more elements of $\mathcal{L}$. Strict and defeasible rules are 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'. As is usual in logic, inference rules can be specified by schemes in which a rule's antecedents and consequent are metavariables ranging over $\mathcal{L}$. Arguments are constructed from a knowledge base, which is assumed to contain three kinds of formulas.

Definition 6 [Knowledge bases] 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 partial preorder 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 and:

- $\mathcal{K}_n$ is a set of (necessary) *axioms*. Intuitively, arguments cannot be attacked on their axiom premises.
- $\mathcal{K}_p$ is a set of *ordinary premises*. Intuitively, arguments can be attacked on their ordinary premises, and whether this results in defeat must be determined by comparing the attacker and the attacked premise (in a way specified below).
- $\mathcal{K}_a$ is a set of *assumptions*. Intuitively, arguments can be attacked on their ordinary assumptions, where these attacks always succeed.

The following definition of arguments is taken from [15], in which for any argument A, the function $\texttt{Prem}$ returns all the formulas of $\mathcal{K}$ (called *premises*) used to build A, $\texttt{Conc}$ returns A's conclusion, $\texttt{Sub}$ returns all of A's sub-arguments, $\texttt{DefRules}$ returns all defeasible rules in A, and $\texttt{TopRule}$ returns the last inference rule used in A.

Definition 7 [Argument] 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. $A_1, \ldots A_n \;\rightarrow\!\!\!/\Rightarrow\; \psi$ 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 $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow\!\!\!/\Rightarrow \psi\}$
   $\texttt{DefRules}(A)$ = $\texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
   $\texttt{TopRule}(A)$ = $\texttt{Conc}(A_1), \ldots \texttt{Conc}(A_n) \rightarrow\!\!\!/\Rightarrow \psi$

Furthermore, $\texttt{DefRules}(A) = \texttt{Rules}(A)/\mathcal{R}_s$. Then 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 $\texttt{Prem}(A) \not\subseteq \mathcal{K}_n$.

The notion of an argument ordering is used in the notion of an argument theory. The argument ordering is a partial preorder $\preceq$ on arguments (with its strict counterpart $\prec$ defined in the usual way), and is assumed to be 'admissible', i.e., firm-and-strict arguments are strictly better than all other arguments, and a strict inference cannot make an argument strictly better or worse than its weakest proper subargument. Note that [13] investigates two example definitions of $\preceq$ in terms of the orderings on $\mathcal{R}_d$ and $\mathcal{K}$.

Definition 8 [Argumentation theories] An *argumentation theory* is a triple $AT = (AS, KB, \preceq)$ where AS is an argumentation system, KB is a knowledge base in AS and $\preceq$ is an admissible ordering on the set of all arguments that can be constructed from KB in AS.

As indicated above, when arguments are inference trees, three syntactic forms of attack are possible: attacking a premise, a conclusion, or an inference. Below these attacks will be called, respectively, undermining, rebutting and undercutting attack. 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, unless indicated otherwise in examples.

Definition 9 [Attacks]
• Argument A *undercuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{B'}$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$.
• Argument 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 φ) iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $\varphi \in \texttt{Prem}(B) \setminus \mathcal{K}_n$. In such a case A *contrary-undermines* B iff $\texttt{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.

Attacks combined with the preferences defined by an argument ordering yield three kinds of defeat. For undercutting attack no preferences will be needed to make it result in defeat, since otherwise a weaker undercutter and its stronger target might be in the same extension. The same holds for the other two ways of attack as far as they involve contraries (i.e., non-symmetric conflict relations between formulas).

Definition 10 [Successful rebuttal, undermining and defeat]
A *successfully rebuts* B if A rebuts B on B' and either A contrary-rebuts B' or $A \not\prec B'$.
A *successfully undermines* B if A undermines B on φ and either A contrary-undermines B or $A \not\prec \varphi$.
A *defeats* B iff A undercuts or successfully rebuts or successfully undermines 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.

In [13], structured argumentation theories are then linked to Dung frameworks:

Definition 11 An *abstract argumentation framework* DF_{AT} *corresponding to an argumentation theory* AT is a pair $\langle \mathcal{A}, Def \rangle$ such that $\mathcal{A}$ is the set of arguments defined by AT as in Definition 7, and Def is the relation on $\mathcal{A}$ given by Definition 10.

Then any semantics for Dung frameworks can be used to define the acceptability status of arguments and their conclusions.

3. Linking Structured Argumentation Theories to Extended Argumentation Frameworks

3.1. Defining Structured Extended Argumentation Frameworks

We build on the previous section's work in order to link structured argumentation theories to a modified version of [7]'s bounded hierarchical *EAF*s. The idea is that the previous section's reference to the argument ordering $\preceq$ is removed; we instead assume a fully abstract partial function $\mathcal{P}$ that extracts orderings from *sets of* arguments that conclude preferences (over other arguments). These sets of preference arguments then *collectively* pref-attack attacks in order to undermine the success of the latter as defeats. In the following section, we then make $\mathcal{P}$ more specific for an argumentation theory that defines $\preceq$ in terms of the two orderings $\leq$ on defeasible rules and $\leq'$ on the knowledge base.

To motivate the generalisation of [7]'s theory to accommodate collective pref-attacks, consider the following informal example argumentation theory in which rules express priorities over other rules (through the use of rule names as in [14]):

Example 12 Let A = $[r_1 :\Rightarrow p, r_2 : p \Rightarrow q]$, B = $[r_3 :\Rightarrow s, r_4 : s \Rightarrow \neg q]$, C_1 = $[r_5 :\Rightarrow r_1 > r_3]$, C_2 = $[r_6 :\Rightarrow r_2 > r_3]$, D_1 = $[r_7 :\Rightarrow r_3 > r_2]$, D_2 = $[r_8 :\Rightarrow r_4 > r_2]$.

A and B attack each other, and A is preferred to B since rule r_3 in B is strictly less than all rules in A, as concluded by arguments C_1 and C_2. Effectively then, it is the arguments C_1 and C_2 that *in combination* express a preference for A over B. In [7] the object level construction of arguments accounts for the conjoining of such arguments C_1 and C_2, so as to obtain a super-argument '$C_1 + C_2$' that attacks the attack from B

to A. This is somewhat inelegant, so that in this paper we conservatively modify [7]'s extended theory to allow for arguments to *collectively* attack attacks, and re-define the notions of defeat, conflict free, and reinstatements sets accordingly. For arbitrary *EAF*s it can be shown that the results in [7] are preserved under this generalisation. In this paper we are interested in *bh-EAF*s, and thus only present collective attacks on attacks (and other modifications) for such *EAF*s:

Definition 13 [*bh-EAFC*]
• A *bh-EAFC* is a tuple ($\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$), where ($\mathcal{A}, \mathcal{C}$) is a *DF* and $\mathcal{D} \subseteq (2^{\mathcal{A}}/\emptyset) \times \mathcal{C}$, and the hierarchical partition of ($\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$) is defined as in Definition 3, replacing a set ϕ of arguments for the single preference argument C.
• A defeats$_S$ B iff $(A, B) \in \mathcal{C}$ and $\neg\exists\phi \subseteq S$ s.t. (ϕ, $(A, B)) \in \mathcal{D}$.
• $S \subseteq \mathcal{A}$ is conflict free iff $\forall A, B \in S$, if $(A, B) \in \mathcal{C}$, then $\exists\phi \subseteq S$ s.t. (ϕ, $(A, B)) \in \mathcal{D}$ (i.e., $\forall A, B \in S, A \nrightarrow^S B$).
• Let $R_S = \{X_1 \rightarrow^S Y_1, \ldots, X_n \rightarrow^S Y_n\}$ where for $i = 1 \ldots n$, $X_i \in S$. R_S is a reinstatement set for $A \rightarrow^S B$, iff $A \rightarrow^S B \in R_S$, and $\forall X \rightarrow^S Y \in R_S$, $\forall\phi$ s.t. (ϕ, $(X, Y)) \in \mathcal{D}$, $\exists X' \rightarrow^S Y' \in R_S$ for some $Y' \in \phi$.
Acceptability and extensions of *bh-EAFC*s are then defined as in Section 2.1.

Two other modifications are worth noting in the above definition. Firstly, we have not included what one would expect to be the following generalisation to the collective case: If (ϕ, (A, B)), (ϕ', $(B, A)) \in \mathcal{D}$, then $\exists Z \in \phi, Z' \in \phi$ s.t. $(Z, Z'), (Z', Z) \in \mathcal{C}$. As will be shown in Section 3.3, this is because when linking structured theories to *bh-EAFC*s one cannot always guarantee that this (or indeed the weaker constraint that an asymmetric attack exists between some Z and Z') follows from the definition of attacks given in Definition 9. The second modification to note is that the definition of conflict free drops the requirement that conflict free sets exclude mutually attacking arguments. We do not want to impose such a constraint at the abstract level; rather we want that it follows from the defined linkage of structured theories to *bh-EAFC*s, that no extension under any of the semantics admits arguments that attack (this will be implied by showing that the linked theories satisfy rationality postulates in Section 3.2). However, it can be shown that despite both these modifications, the key results for the extended theory defined in Definition 13 still hold (proofs of all the results in this paper can be found in [11]):

Proposition 14 [Fundamental lemma and Monotonicity of Characteristic Function]
Let Δ = ($\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$) be a *bh-EAFC*. Then:
1) If S is an admissible extension of Δ, and A, A' arguments acceptable w.r.t S, then: S' = $S \cup \{A\}$ is admissible; A' is acceptable w.r.t. S'.
2) Let S and S' be conflict free subsets of $\mathcal{A}$ such that $S \subseteq S'$. Then $\mathcal{F}(S) \subseteq \mathcal{F}(S')$.[3]

We are now ready to link structured theories to *bh-EAFC*s.

[3]This result is to be expected given that the requirements that contradictory preference arguments symmetrically attack, and that conflict free sets exclude symmetrically attacking arguments, are only required to show (in [7]) that iterating $\mathcal{F}$ from the empty set yields a fixed point and so defines the grounded extension for *arbitrary EAF*s. For *bh-EAFC*s, it follows from 1) that all admissible extensions form a complete partial order w.r.t. set inclusion, and 2) guarantees the existence of a least fixed point for $\mathcal{F}$ that identifies a finitary *bh-EAFC*'s grounded extension.

Definition 15 [Extended Argumentation Theory, Arguments and Preference Function]
• An *extended argumentation system* is a triple $EAS = (\mathcal{L}, ^-, \mathcal{R})$
• An *extended knowledge base* is a set $EKB = \mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a$
• An *extended argumentation theory* is a tuple $EAT = (EAS, EKB)$
• Let $\mathcal{A}$ denote the set of arguments defined by EAT as in Definition 7. We say that $\mathcal{P}$ is a partial function defined by EAT, where:

$$\mathcal{P} : X \longrightarrow Pow(\mathcal{A} \times \mathcal{A}) \text{ (for some } X \in \mathcal{A}).$$

When instantiating a *bh-EAFC*, we note that since A may rebut or undermine B on more than one sub-argument, respectively premise, then by Definition 10, A does not defeat B if A does not contrary-rebut/undermine B, and *for all* rebutted sub-arguments B' and undermined premises ϕ of B, $A \prec B'$ and $A \preceq \phi$. This will be made explicit when defining attacks on attacks in the following definition.

Definition 16 [*bh-EAFC* for structured arguments] A *bh-EAFC*$_{EAT}$ corresponding to an EAT, henceforth referred to as a *structured EAF*, is a *bh-EAFC* $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ such that:

1. $\mathcal{A}$ is the set of arguments defined by EAT as in Definition 7;
2. $(A, B) \in \mathcal{C}$ iff A undercuts, rebuts or undermines B according to Definition 9;
3. $(\phi, (A, B)) \in \mathcal{D}$ iff $(A, B) \in \mathcal{C}$, and:
   (a) $\forall B' \in \mathtt{Sub}(B)$ s.t. A rebuts or undermines B on B', $\exists \phi' \subseteq \phi$ s.t. $A \prec B' \in \mathcal{P}(\phi')$, and ϕ is a minimal (under set inclusion) set satisfying this condition.
   (b) A does not contrary undermine, contrary rebut or undercut B (since by Definition 10 these attacks succeed as defeats irrespective of preferences).
   (c) it is not the case that A is firm and strict and B is plausible or defeasible (since by the admissibility of argument orderings described prior to Definition 8, it must be that $B \prec A$).

We say that E is an extension of an EAT iff E is an extension of *bh-EAFC*$_{EAT}$.

3.2. Satisfaction of Rationality Postulates by Structured EAFs

In [13], DF_{AT}s are shown to satisfy [5]'s rationality postulates. Structured *EAF*s also satisfy these rationality postulates. Firstly, the sub-argument closure and closure under strict rules postulates are unconditionally satisfied:

Theorem 17 [Sub-argument Closure] Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$ and E any of its extensions under a given semantics subsumed by complete semantics. Then for all $A \in E$: if $A' \in \mathtt{Sub}(A)$ then $A' \in E$.

Theorem 18 [Closure under strict rules] Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$ and E any of its extensions under a given semantics subsumed by complete semantics. Then $\{\mathtt{Conc}(A) | A \in E\} = Cl_{\mathcal{R}_s}(\{\mathtt{Conc}(A) | A \in E\})$ [4].

In [13] it is shown that DF_{AT}s satisfy the consistency postulates under a number of assumptions that are more fully described in [13]:

[4] $Cl_{\mathcal{R}_s}(P)$, where $P \subseteq \mathcal{L}$ is the smallest set containing P and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{\mathcal{R}_s}(P)$

(Ass1) the argumentation system's strict rules are closed under 'transposition'[5].
(Ass2) the closure of $\mathcal{K}_n$ under strict rule application is consistent.
(Ass3) the argumentation theory is 'well-formed'.
(Ass4) the argument ordering is 'reasonable'.

In this paper we refer to assumptions **Ass1-3** straightforwardly applied to the extended argumentation theories of Definition 15. We discuss **Ass4** after first describing an assumption that essentially expresses (at the level of the instantiating EAT) an analogue of the omitted constraint on contradictory sets of preference arguments discussed earlier:

Definition 19 [**Ass5**] Let $\Delta = (\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$, and suppose $\phi, \psi \subseteq \mathcal{A}$ s.t. $B \prec A \in \mathcal{P}(\phi)$, $A \prec B \in \mathcal{P}(\psi)$. Then Δ satisfies **Ass5** if for some $X \in \phi$, $Y \in \psi$, either X and Y have contradictory conclusions, or there exists some set of strict rules extending X to the argument $X+$ s.t. $X+$ and Y have contradictory conclusions.

We informally illustrate **Ass5** with Example 12, in which $(\{C_1, C_2\}, (B, A)) \in \mathcal{D}$ and $(\{D_1, D_2\}, (A, B)) \in \mathcal{D}$. Assume the strict rules contain the axioms of a partial order, including the rule for asymmetry: $o_4 : X > Y \rightarrow \neg(Y > X)$, where X and Y range over rule names. Then D_1 can be extended to D_1' = $[r_7 :\Rightarrow r_3 > r_2, o_4 : r_3 > r_2 \rightarrow \neg(r_2 > r_3)]$ whose conclusion contradicts C_2's conclusion. Hence D_1' asymmetrically attacks C_2. Before discussing **Ass4**, we recall some notation from [13]:

Notation 20 $M(B)$ denotes the maximal fallible sub-arguments 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).

Ass4's reasonable ordering assumption captures the intuition that given arguments A and B, both of which are plausible or defeasible and such that $B \prec A$, then there must be some $B' \in M(B)$ such that:

i B' is not stronger than any maximal fallible sub-argument in $M(B)$ (i.e., $M(B)$ contains a $\preceq$ minimal element);
ii) $B' \prec A$ (since otherwise it cannot be that $B \prec A$ given that B consists of $M(B)$ extended by strict rules that by the admissibility of $\preceq$ cannot weaken the arguments in $M(B)$)

Articulating a counterpart to the **Ass4** in the context of *structured EAF*s, recall that we are interested in cases where $(\phi, (B, A)) \in \mathcal{D}$, where for each sub-argument A' of A rebutted or undermined by B, there is a subset of ϕ that expresses a preference for A' over B. Also, since contradictory preferences can be expressed, and so the existence of $\preceq$ minimal arguments cannot be guaranteed, we also need to express the assumption in the context of some set of arguments E in which such a minimal argument does exist:

Definition 21 [**Ass6**] Let $\Delta = (\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$, $E \subseteq \mathcal{A}$, $A, B \in E$, and $(\phi, (B, A)) \in \mathcal{D}$. Let there exist at least one argument $X \in M(B)$ that is a $\preceq$ minimal argument in E in the sense that:

$$\text{for all } B'' \in M(B), \neg\exists \psi \subseteq E \text{ s.t. } B'' \prec X \in \mathcal{P}(\psi).$$

Then Δ satisfies **Ass6** if $\forall A'$ s.t. A' is a sub-argument of A and B rebuts or undermines A on A', $\exists B' \in M(B)$ that is $\preceq$ minimal in E, $\exists \phi' \subseteq \phi$ s.t. $B' \prec A' \in \mathcal{P}(\phi')$

[5] i.e., $s = \varphi_1, \ldots, \varphi_n \rightarrow \psi \in \mathcal{R}_s$ iff for $i = 1 \ldots n$, $\varphi_1, \ldots, \varphi_{i-1}, -\psi, \varphi_{i+1}, \ldots, \varphi_n \rightarrow -\varphi_i \in \mathcal{R}_s$

Notice that if for a finite $M(B)$ there is no argument in $M(B)$ that is $\preceq$ minimal in E, then $\forall B' \in M(B)$, $\exists B'' \in M(B)$, $\exists \phi \subseteq E$ s.t. $B'' \prec B' \in \mathcal{P}(\phi)$. This in turn implies that for some $B', B'' \in M(B)$, $\exists \phi, \psi \subseteq E$ s.t. $B' \prec B'' \in \mathcal{P}(\phi)$, $B'' \prec B' \in \mathcal{P}(\psi)$. Then by **Ass5**, there must be some $X \in \phi$, $Y \in \psi$ such that Y, and X or $X+$ extending X with strict rules, have contradictory conclusions. Hence **Ass5** effectively implies that a $\preceq$ minimal $X \in M(B)$ exists in a set free of arguments with contradictory conclusions. We can now state the following theorems:

Theorem 22 Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$ satisfying **Ass1-3**, **Ass5**, and **Ass6**. Let E be any of its extensions under a given semantics subsumed by complete semantics. Then $\{\texttt{Conc}(A) | A \in E\}$ is consistent.

Theorem 23 Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$ satisfying **Ass1-3**, **Ass5**, and **Ass6**. Let E be any of its extensions under a given semantics subsumed by complete semantics. Then $Cl_{\mathcal{R}_s}(\{\texttt{Conc}(A) | A \in E\})$ is consistent.

3.3. An Example Extended Argumentation Theory

In this section we describe an extended argumentation theory (*EAT*) and its *structured EAF*. As in Example 12 we assume arguments constructed from named rules that may express priorities over other rules. We also assume that any *EAT* contains (in its component *EAS*) strict rules axiomatising a partial order (x,y,z are meta-variables ranging over rule names and $o2$ and $o3$ are the transpositions of $o1$):

- $o_1 : (y > x) \wedge (z > y) \rightarrow (z > x)$
- $o_2 : (y > x) \wedge \neg(z > x) \rightarrow \neg(z > y)$
- $o_3 : (z > y) \wedge \neg(z > x) \rightarrow \neg(y > x)$
- $o_4 : (y > x) \rightarrow \neg(x > y)$

We then assume that $B \prec A \in \mathcal{P}(\phi)$ if the arguments in ϕ conclude rule priorities such that A is stronger than B under the last link principle [2]:

Definition 24 [Conclusion of $\prec_s$ by a set of arguments] Let $\Gamma = \{r_1 : l_1, \ldots, r_n : l_n\}$ be a set of objects named by wff of $\mathcal{L}$, and $\geq$ a partial ordering on Γ (with its strict counterpart $>$ defined in the usual way). Let $\Gamma' \subseteq \Gamma, \Gamma'' \subseteq \Gamma$. Then for some set ϕ of arguments:

ϕ is said to conclude that $\Gamma' \prec_s \Gamma''$, iff $\exists r_i : l_i \in \Gamma'$ s.t. $\forall r : l \in \Gamma''$, $r > r_i$ is the conclusion of an argument in ϕ.

Definition 25 [$\mathcal{P}$ defined under the last link principle] Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$, $A, B \in \mathcal{A}$, $\phi \subseteq \mathcal{A}$. Then $B \prec A \in \mathcal{P}(\phi)$ under the last link principle iff

1. ϕ concludes $\texttt{LastDefRules}(B) \prec_s \texttt{LastDefRules}(A)$; or
2. $\texttt{LastDefRules}(B)$ and $\texttt{LastDefRules}(A)$ are empty and ϕ concludes $\texttt{Prem}(B) \prec_s \texttt{Prem}(A)$

Let us now illustrate how a *structured EAF* is instantiated by arguments constructed from an *EAT*. For simplicity, our example is with domain-specific inference rules, mostly with empty antecedents. Consider the following defeasible rules:

r_1:	$\Rightarrow a$	p_1:	$\Rightarrow r_3 > r_1$	m_1:	$\Rightarrow p_3 > p_1$
r_2:	$\Rightarrow \neg a$	p_2:	$\Rightarrow r_2 > r_3$	m_2:	$\Rightarrow p_1 \approx p_2$
r_3:	$\Rightarrow b$	p_3:	$\Rightarrow r_1 > r_2$		

Indexing the inferences with the names of the rules applied, we have the mutually rebutting arguments $X_1 : \Rightarrow_{r1} a$, $X_2 : \Rightarrow_{r2} \neg a$, and:

A_1:	$\Rightarrow_{p1} r_3 > r_1$	B_1:	$\Rightarrow_{p3} r_1 > r_2$	C:	$\Rightarrow_{m1} p_1 < p_3$
A_2:	$\Rightarrow_{p2} r_2 > r_3$	B_2:	$B_1 \rightarrow_{o4} \neg(r_2 > r_1)$	D:	$\Rightarrow_{m2} p_1 \approx p_2$
A_3:	$A_1, A_2 \rightarrow_{o1} r_2 > r_1$				

Applying the last link principle, $\mathcal{P}(\{A_3\}) = \{X_1 \prec X_2\}$ hence $(\{A_3\}, (X_1, X_2)) \in \mathcal{D}$, and $\mathcal{P}(\{B_1\}) = \{X_2 \prec X_1\}$ hence $(\{B_1\}, (X_2, X_1)) \in \mathcal{D}$, as illustrated in Figure 1.

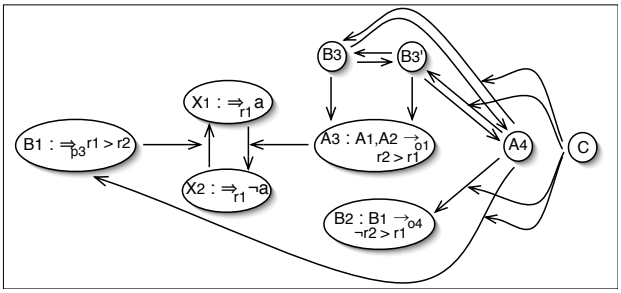

Figure 1. *Structured EAF* for example EAT

Now note that B_1 and A_3 do not attack each other. Furthermore, although B_1 can be extended with a strict rule to B_2, with A_3 and B_2 having contradictory conclusions (illustrating satisfaction of **Ass5**), A_3 and B_2 do not attack each other since both have a strict top rule. However, with transpositions of these strict top rules, both can be extended to attack the other on a defeasible subargument:

A_4:	$A_3 \rightarrow_{o4} \neg(r_1 > r_2)$	(rebutting B_1, and so B_2, B_3 and B_3' on B_1)
B_3:	$B_2, A_1 \rightarrow_{o2} \neg(r_2 > r_3)$	(rebutting A_2, and so A_3 and A_4 on A_2)
B_3':	$B_2, A_2 \rightarrow_{o3} \neg(r_3 > r_1)$	(rebutting A_1, and so A_3 and A_4 on A_1)

Now, $\texttt{LastDefRules}(A_4) = \{p_1, p_2\}$, $\texttt{LastDefRules}(B_1) = \{p_3\}$, and for $i = 1, 2, 3, 3'$, $A_4 \prec B_i \in \mathcal{P}(\{C\})$ and so $(\{C\}, (A_4, B_i)) \in \mathcal{D}$. Also, $\texttt{LastDefRules}(B_3) = \{p_1, p_3\}$ and $\texttt{LastDefRules}(A_2) = \{p_2\}$. Then it is easy to verify that no ϕ pref-attacks B_3's attack on A_2. Similarly, $\texttt{LastDefRules}(B_3') = \{p_2, p_3\}$ and $\texttt{LastDefRules}(A_2) = \{p_1\}$, so no ϕ pref-attacks B_3's attack on A_2. This means that A_3 will not be in any extension: in grounded semantics this is since neither A_1 nor A_2 is in the grounded extension, while in the other semantics this is since each extension contains either A_1 or A_2 but not both (since each extension contains B_3 or B_3'). So in all extensions X_1's attack on X_2 is successful. Since C and D are not attacked, they will be in all extensions. Hence all attacks from A_4 on B_1, B_2, B_3 and $B3'$ are attacked by C, so that each extension will contain B_1, B_2 and B_3 or B_3'. Hence, in no extension is X_2's attack on X_1 successful, and so for any E under any of [6]'s semantics, $X_1 \rightarrow^E X_2$ but not $X_2 \rightarrow^E X_1$. So all such extensions contain X_1 but not X_2.

Finally, for the *structured EAF*s obtained by the instantiating EATs in this section, the theorems in the previous section imply that all the rationality postulates hold, given that we can show that **Ass5** and **Ass6** hold under the last link principle:

Proposition 26 Let $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ be a *bh-EAFC*$_{EAT}$, where the strict rules in EAT include $o1 \ldots o4$. Let $\mathcal{P}$ be defined under the last link principle. Then $(\mathcal{A}, \mathcal{C}, \mathcal{D})$ satisfies **Ass5** and **Ass6**.

4. Conclusions

In this paper we have presented an abstract formalism for reasoning about preferences in structured extended argumentation frameworks. We motivated modifications to [7]'s extended argumentation, dropping [7]'s constraints on conflict free sets and (sets of) arguments expressing contradictory preferences, and enabling collective pref-attacks on attacks. We showed that the fundamental results that hold for bounded hierarchical *EAF*s also hold for the modified theory. We then instantiated the modified *EAF*s with [13]'s structured argumentation theories modified so as to allow for sets of arguments to express preferences over other arguments. We then showed that the obtained instantiated *structured EAF*s satisfy [5]'s closure and consistency postulates, and described an instantiation by arguments built from rules that can express priorities over other rules. The abstract specification of the instantiating structured argumentation theories means that our work enables principled development of novel and existing systems. In future work we will thus investigate how various existing argumentation systems (e.g. [4]'s assumption based argumentation) that are shown to be a special case of [13]'s instantiaton of *DF*s, can now be extended in a principled way to enable argumentation based reasoning about preferences over other arguments. Future work will also investigate the more general case of non-hierarchical *EAF*s, and application of preference criteria other than the last link principle (such as the weakest link principle).

References

[1] 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).

[2] L. Amgoud, L. Bodenstaff, 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, ASPIC IST-FP6-002307, (2006).

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

[4] 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).

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

[6] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence*, **77**, 321–357, (1995).

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

[8] T. J. M. Bench-Capon and S. Modgil. Case law in extended argumentation frameworks. In *ICAIL*, 118–127, (2009).

[9] S. Modgil. An argumentation based semantics for agent reasoning. In *Proc. Workshop on Languages, methodologies and development tools for multi-agent systems (LADS 07)*, 37–53, UK, (2007).

[10] S. Modgil and M. Luck. Argumentation based resolution of conflicts between desires and normative goals. In *Proc. 5th Int. Workshop on Argumentation in Multi-Agent Systems*, 252–263, (2008).

[11] S. Modgil and H. Prakken. Technical Report: Proofs for *Structured EAF*s. In http://people.cs.uu.nl/henry/ASPIC-EAF-TR.pdf, (2010).

[12] J.L. Pollock. Justification and defeat. *Artificial Intelligence*, **67**, 377–408, (1994).

[13] H. Prakken. An abstract framework for argumentation with structured arguments. To appear in: *Argument and Computation*, **1**, (2010). www.cs.uu.nl/research/techreps/UU-CS-2009-019.html

[14] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, **7**, 25–75, (1997).

[15] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, **90**, 225–279, (1997).