Reasoning About Preferences in Argumentation Frameworks TR-09-01

Sanjay Modgil[a*]
[a]Department of Computer Science, Kings College London
Strand, London WC2R 2LS

Abstract

The abstract nature of Dung's seminal theory of argumentation accounts for its widespread application as a general framework for various species of non-monotonic reasoning, and, more generally, reasoning in the presence of conflict. A Dung argumentation framework is instantiated by arguments and a binary conflict based attack relation, defined by some underlying logical theory. The justified arguments under different extensional semantics are then evaluated, and the claims of these arguments define the inferences of the underlying theory. To determine a unique set of justified arguments often requires a preference relation on arguments to determine the success of attacks between arguments. However, preference information is often itself defeasible, conflicting and so subject to argumentation. Hence, in this paper we extend Dung's theory to accommodate arguments that claim preferences between other arguments, thus incorporating meta-level argumentation based reasoning about preferences in the object level. We then define and study application of the full range of Dung's extensional semantics to the extended framework, and study special classes of the extended framework. The extended theory preserves the abstract nature of Dung's approach, thus aiming at a general framework for non-monotonic formalisms that accommodate defeasible reasoning *about* as well as with preference information. We illustrate by formalising argument based logic programming with defeasible priorities in the extended theory.

Keywords: Argumentation, Dung, Preferences, Non-monotonic Reasoning, Logic Programming

1 Introduction

1.1 Background

The formal study of argumentation has come to be increasingly central as a core study within Artificial Intelligence [12]. Logic based models of argumentation are being applied to formalisation of defeasible reasoning and conflict resolution over beliefs and goals, and to decision making over actions [1, 8, 24, 25, 35]. The inherently dialectical

*Email address: sanjay.modgil@kcl.ac.uk. Phone: +44 (0)207 8481631

nature of these models have provided foundations for formalisation of argumentation-based dialogues [5], where, for example, one agent seeks to persuade another to adopt a belief it does not already hold to be true [33], or when agents deliberate about what actions to execute [21], or negotiate over resources [4]. Furthermore, recent major research projects [6, 7] are developing general models of argumentation based inference, decision making and dialogue, and implementations of these models for deployment in agent and semantic grid applications.

Many of the above theoretical and practical developments build on Dung's seminal theory of argumentation [19]. A Dung *argumentation framework* is a directed graph consisting of a set of arguments $Args$ and a binary conflict based *attack* relation $\mathcal{R}$ on $Args$. The extensions, and so justified arguments of a framework are then defined under different semantics, where the choice of semantics equates with varying degrees of scepticism or credulousness. Extensions are defined through application of an 'acceptability calculus', whereby an argument $X \in Args$ is said to be *acceptable* with respect to $S \subseteq Args$ iff any argument Y that attacks X is itself attacked by some argument Z in S (intuitively, any such Z is said to *reinstate* X). For example, if S is a maximal (under set inclusion) set such that all its contained arguments are acceptable with respect to S, then S is said to be an extension under the *preferred* semantics.

Recent years have witnessed intensive formal study, development and application of Dungs ideas in various directions. This can be attributed to the abstract nature of a Dung argumentation framework, and the encoding of intuitive, generic principles of commonsense reasoning through application of the acceptability calculus. The underlying logic, and definition of the logic's constructed arguments $Args$ and attack relation $\mathcal{R}$ is left unspecified, thus enabling instantiation of a framework by various logical formalisms. A theory's inferences can then be defined in terms of the claims of the justified arguments constructed from the theory (an argument essentially being a proof of a candidate inference - the argument's claim - in the underlying logic). Dung's theory can therefore be understood as a *semantics* for non-monotonic reasoning. In this view, what appropriately accounts for the correctness of an inference is that it can be shown to rationally prevail in the face of opposing inferences, where it is application of the acceptability calculus that encodes logic neutral, rational means for establishing such standards of correctness. Indeed, many logic programming formalisms and non-monotonic logics (e.g. default, auto-epistemic, defeasible, non-monotonic modal logics and certain instances of circumscription) have been shown to conform to Dung's semantics [13, 18, 19, 20].

Dung's extensional semantics may yield multiple extensions. The *sceptically* justified arguments are those that appear in every extension. However, one may then be faced with the problem of how to choose between conflicting *credulously* justified arguments that belong to at least one, but not all extensions. To illustrate, consider two individuals **P** and **Q** exchanging arguments A, B ... about the weather forecast:

$\mathbf{P}_1$: "Today will be dry in London since the BBC forecast sunshine" = A

$\mathbf{Q}_1$: "Today will be wet in London since CNN forecast rain" = B

A and B claim contradictory conclusions and so attack each other (symmetrically attack). Under Dung's *preferred* semantics, there are two extensions: $\{A\}$ and $\{B\}$. Neither argument is sceptically justified. One solution is to provide some means for preferring one argument to another. Some works (e.g., [32, 34]) formalise the role of preferences in the underlying logical formalisms that instantiate a Dung framework. For example, in [32], if X undercuts Y (where 'undercut' denotes a certain type of conflict based interaction), then $X\mathcal{R}Y$ if Y is not stronger than (preferred to) X. Other approaches formalise the role of preferences at an abstract level. In Amgoud's Preference based Argumentation Frameworks [3], a Dung framework is augmented with a preference ordering on $Args$, so that an attack by X on Y is successful only if Y is not preferred to X. In Bench-Capon's Value based Argumentation Frameworks (*VAF*s) [11], a Dung framework is augmented with values and value orderings, so that an attack by X on on Y is successful only if the *value* promoted by Y is not ranked higher than the value promoted by X according to a given ordering on values.

Examining the role of preferences in the above weather example, one might reason that A is preferred to B because the BBC are deemed more trustworthy than CNN. Hence B does not successfully attack A, and so this attack can be removed and we are only left with the successful attack from A to B. Thus only $\{A\}$ will be a preferred extension; A is sceptically justified. This example illustrates *resolution* of an argumentation framework obtained by replacing symmetric attacks with asymmetric attacks. Properties relating frameworks and their resolutions have been studied in [9] and [26]. Properties of frameworks obtained through removal of asymmetric attacks have also been studied in the context of *VAF*s.

1.2 Overview of Paper

Information required to determine the success of an attack is often assumed pre-specified, as a given preference or value ordering. However, preference information may be contradictory, given that preferences may vary according to context, and distinct sources may valuate the strengths of arguments using different criteria, or indeed assign different valuations for any given criterion. Thus, one often needs to reason, and indeed argue *about*, as well as *with*, defeasible and possibly conflicting preference information. To illustrate, consider continuation of the above dialogue about the weather:

$\mathbf{P}_2$: "But the BBC are more trustworthy than CNN" = C
$\mathbf{Q}_2$: "However, statistically CNN are more accurate forecasters than the BBC" = C'
$\mathbf{Q}_3$: "And basing a comparison on statistics is more rigorous and rational than basing a comparison on your instincts about their relative trustworthiness" = E.

Argument C is not an argument that attacks B; rather it is an argument expressing a preference for A over B. However, C' expresses a preference for B over A. C and C' attack each other since they express contradictory preferences. Finally, argument E claims that C' is preferred to C, and so only C' successfully attacks C. We thus have a sceptically justified argument C' claiming that B is preferred to A, and so B is sceptically justified.

In this paper we are thus motivated to extend Dung's theory to accommodate argumentation *about* preferences. The paper is organised as follows. In Section 2 we review Dung's theory of argumentation. In the sections that follow we describe our contributions to the formal study and application of argumentation:

• In sections 3 and 4, Dung's theory of argumentation is extended to integrate 'metalevel' argumentation about preferences between arguments. The extended theory preserves the abstract nature of Dung's approach. Arguments expressing preferences (*preference arguments*) are simply nodes in a graph, and application of preferences is abstractly characterised; by defining a new attack relation that originates from a preference argument, and that *attacks an attack* between the arguments that are the subject of the preference claim. Thus, no commitment is made to how preferences are defined and applied in the formalism instantiating the framework. Preference arguments expressing contradictory preferences attack each other, where these attacks can then themselves be attacked by preference arguments. We then extend Dung's acceptability calculus so that both arguments *and* attacks need to be reinstated. The extensional semantics for the extended framework are then defined in exactly the same way as for a Dung framework, and some basic results that hold of the extensional semantics for a Dung framework are also shown to hold for the extended framework. Our aim is to therefore lay foundations for straightforward modification of applications and developments of Dung's work to accommodate argumentation about preferences.

• The extended theory provides a general setting for formally studying the application of preferences in argumentation, and, more generally, the removal of attacks. Thus, the above mentioned work on resolutions of frameworks [9, 26] can now be extended to account for the use of preference arguments in obtaining such resolutions. In this paper we study two special classes of extended argumentation framework: Section 5 studies *hierarchical* frameworks that posit restrictions on interactions between arguments at the object and metalevel, and Section 6 studies *preference symmetric* frameworks that only allow for removal of attacks between symmetrically attacking arguments.

• The extended theory is proposed as a unifying framework in which to formalise and extend works that augment Dung's framework with preferences [3, 11]. In support of this proposal, Section 5 describes how value based argumentation [11] can be formalised and extended in hierarchical frameworks in order to integrate metalevel reasoning *about* values and value orderings.

• The extended theory is proposed as a semantics for object level non-monotonic formalisms that accommodate defeasible reasoning about priorities. Such formalisms extend the object level logical languages for argument construction with rules for deriving priorities amongst rules, e.g., in default logic [14] and logic programming [24, 34]. One can then construct 'priority arguments' whose claims determine preferences between other mutually attacking arguments. Arguments claiming conflicting priorities may be constructed and preferences between these can be established on the basis of other priority arguments. Our approach generalises these formalisms in that preferences may be based not only on rule priorities, but on any criterion for valuating argument strength, including criteria that relate to the argument as a whole (such as the value promoted by an argument). Furthermore, these formalisms do not straightforwardly allow for application of Dung's acceptability calculus, and evaluation of the justified arguments under the full range of Dung's semantics. The extended framework provides for such evaluation, through provision of an abstract characterisation of preference application while retaining the basic conceptual machinery of Dung. In Section 7 we show how the extended theory can serve as a semantics for Prakken and Sartor's argument based logic programming with defeasible priorities (ALP-DP) [34][1]. Arguments defined by an ALP-DP theory instantiate a preference symmetric framework, so that in contrast with [34], one can then evaluate the justified arguments of an ALP-DP theory under *all* of Dung's semantics.

In Section 8 we discuss related work; in particular the work of Kakas and Moraitis

[1] Note that extended frameworks and their instantiation by [34]'s logic programming formalism were first introduced in [28]. This paper contains a modified (simplified) account of the formalism described in [28], as well as additionally providing proofs of results, a formal description of the special cases of the extended frameworks and their properties, and formalisation of metalevel reasoning about values and value orderings in the extended framework.

[23, 24], and conclude and indicate directions for future work in Section 9. Finally, the reader is referred to the appendix for proofs of the paper's propositions.

2 Dung's Theory of Argumentation

In this Section we review Dung's theory of argumentation [19].

Definition 1 A Dung argumentation framework is a tuple $AF = (Args, \mathcal{R})$, where $Args$ is a set of arguments, and $\mathcal{R} \subseteq Args \times Args$.
Figure 1 shows an argumentation framework in which an arrow from X to Y denotes that $(X, Y) \in \mathcal{R}$.

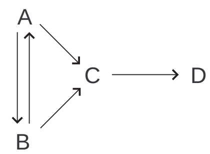


Figure 1: A Dung argumentation framework

Dung then defines the acceptability of arguments, and the characteristic function and admissible extensions of a framework.

Definition 2 Let $AF = (Args, \mathcal{R})$, $S \subseteq Args$ and let $A, B, C, \ldots$ denote arguments in $Args$. Then:

1. S is conflict free iff $\forall A, B \in S$ it is not the case that $(A, B) \in \mathcal{R}$
2. A is acceptable with respect to S iff $\forall B \in Args$: if $(B, A) \in \mathcal{R}$ then there is a $C \in S$ such that $(C, B) \in \mathcal{R}$
3. The characteristic function of AF, denoted F_{AF}, is defined as follows:
   - $F_{AF} : 2^{Args} \mapsto 2^{Args}$
   - $F_{AF}(S) = \{A | A$ is acceptable w.r.t. $S\}$
4. If S is conflict free, then S is an *admissible* extension of AF iff each argument in S is acceptable with respect to S (i.e., $S \subseteq F_{AF}(S)$)

Intuitively, an argument A is acceptable with respect to S if for any argument B that attacks A, there is a C in S that attacks B, in which case C is said to defend or 'reinstate' A. An admissible extension S can then be interpreted as a coherent defendable position. From hereon, since we will refer to an arbitrary but fixed argumentation framework, we will write F instead of F_{AF}.

Admissibility is augmented by preferred, complete, stable and grounded semantics:

Definition 3 Let $AF = (Args, \mathcal{R})$, S a conflict free subset of $Args$, and F the characteristic function of AF. Then:

- S is a preferred extension iff S is a set inclusion maximal admissible extension

- S is a complete extension iff each argument which is acceptable w.r.t. S is in S ($S = F(S)$)
- S is a stable extension iff $\forall B \notin S$, $\exists A \in S$ such that $(A, B) \in \mathcal{R}$
- S is the grounded extension iff S is the least fixed point of F (the smallest complete extension).

For $s \in \{$complete, preferred, stable, grounded$\}$, an argument is said to be sceptically justified under the s semantics if it belongs to all extensions. An argument is said to be only credulously justified under the s semantics if it belongs to at least one, but not all, extensions. Sceptical and credulous justification coincide for the grounded semantics given that a framework only ever has a single grounded extension.

The admissible extensions of the framework in Figure 1 are: $\emptyset$, $\{A\}$, $\{B\}$, $\{A, D\}$, and $\{B, D\}$. $\{A, D\}$, and $\{B, D\}$ are the preferred and stable extensions. $\emptyset$, $\{A, D\}$, and $\{B, D\}$ are the complete extensions, and $\emptyset$ the grounded extension. Note that D is sceptically justified under the preferred semantics, but is not sceptically justified under the grounded semantics.

3 Extended Argumentation Frameworks

In this section we motivate extending Dung's framework to include arguments that express preferences between other arguments, and so determine whether attacks succeed. Following convention, we may in what follows synonymously write 'defeats' instead of 'successfully attacks'.

Recall the weather forecast example in Section 1 in which arguments A and B symmetrically attack ((A, B), $(B, A) \in \mathcal{R}$). Hence $\{A\}$ and $\{B\}$ are admissible. We have that A is preferred to B (claimed by argument C). Hence B does not successfully attack (defeat) A, but A does defeat B. If we were to evaluate admissibility on the basis of defeats, then $\{A\}$ is admissible, and $\{B\}$ is not an admissible extension.

The impact of argument C could conceivably be modelled by letting C attack B ($\Delta 1$ in Figure 2). This would yield the required result: only $\{A\}$ is admissible. However, suppose an argument D that defeats A (e.g. D = "the BBC forecast is for Glasgow and not London"). Then $\{B\}$ would still not be admissible ($\Delta 2$ in Figure 2), which is clearly inappropriate. C expresses a preference for A over B, but if A is defeated by another argument, then B should be reinstated.

Intuitively, C is an argument about the relationship between A and B. Specifically, in expressing a preference for A over B, C is an argument for A's defence against B's attack on A. That is, C *defence* attacks ***B's attack on** A* ($\Delta 3$ in Figure 2) so that B's attack on A does not succeed as a defeat. B's attack on A is, as it were, cancelled out, and we are left with A defeating B. Now, if D defeats A, D reinstates B and $\{B, D\}$ is an admissible extension of $\Delta 4$ in Figure 2.

Of course, given C' claiming a preference for B over A and so *defence* (*d*) attacking A's attack on B, then the issue of which attack succeeds as a defeat is once again unresolved. Intuitively, C and C' claim contradictory preferences and so attack each other ($\Delta 5$ in Figure 2). These attacks can themselves be subject to *d* attacks in order to determine the defeat relation between C and C' and so A and B. In $\Delta 6$ in Figure 2, E *d* attacks the attack from C to C', and so determines that C' defeats C. Hence, C's *d*-attack on B's attack on A is cancelled out, and we are left with B defeating A; the

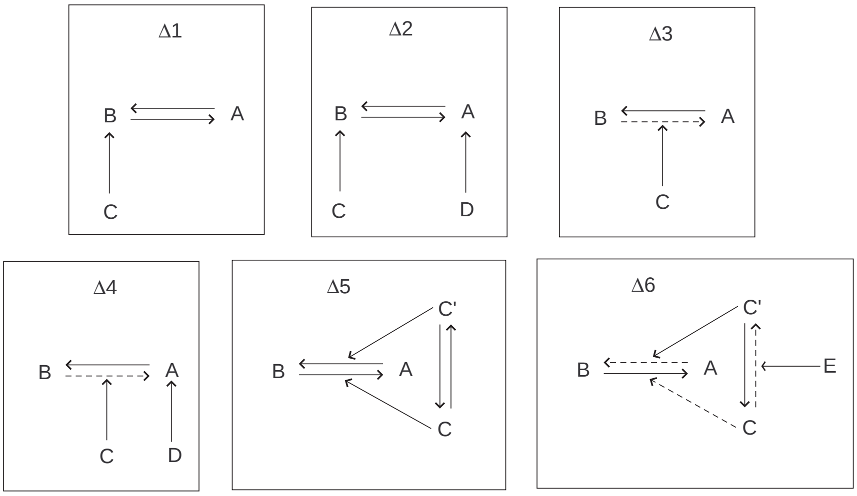


Figure 2: Arguing about the weather

discussion concludes in favour of **Q**'s argument that it will be a wet day in London.

We now formally define the elements of an *Extended Argumentation Framework* (*EAF*). As suggested by the above example, this involves extending Dung's argumentation framework to include a second *attack* relation $\mathcal{D}$ that ranges from arguments X to attacks $(Y, Z) \in \mathcal{R}$, where $\mathcal{R}$ is the standard binary attack relation in a Dung framework. If X attacks (Y, Z) then X expresses that Z is preferred to Y. If X' attacks (Z, Y), then X' expresses that Y is preferred to Z. Hence, *EAF*s are required to conform to the constraint that any such arguments expressing contradictory preferences must attack each other (i.e., (X, X'), $(X', X) \in \mathcal{R}$).

Definition 4 An *Extended Argumentation Framework* (*EAF*) is a tuple $(Args, \mathcal{R}, \mathcal{D})$ such that $Args$ is a set of arguments, and:

- $\mathcal{R} \subseteq Args \times Args$
- $\mathcal{D} \subseteq Args \times \mathcal{R}$
- If $(X,(Y, Z))$, $(X',(Z, Y)) \in \mathcal{D}$ then (X, X'), $(X', X) \in \mathcal{R}$

Notation 1 From hereon we may use the following notation:

- $Y \rightharpoonup Z$ denotes $(Y, Z) \in \mathcal{R}$. If in addition $(Z, Y) \in \mathcal{R}$, we may write $Y \rightleftharpoons Z$.
- $X \twoheadrightarrow (Y \rightharpoonup Z)$ denotes $(X,(Y, Z)) \in \mathcal{D}$.

Example 1 The *EAF* for the weather forecasting example is:
$Args = \{A, B, C, C', E\}$
$\mathcal{R} = \{(A, B), (B, A), (C, C'), (C', C)\}$
$\mathcal{D} = \{(C, (B, A)), (C', (A, B)), (E, (C, C'))\}$

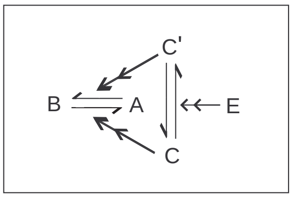


Figure 3: The *EAF* for the weather forecast example

In an *EAF*, preferences are not defined by some externally given preference ordering, but are themselves claimed by arguments. Intuitively, given that an argument A attacks B, then one would reason that A defeats B, only if the arguments S that one is currently committed to, contain no argument claiming that B is preferred to A. That is to say, the success of an attack as a defeat is parameterised w.r.t. the preference arguments available in some such set S of arguments[2].

Definition 5 Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an *EAF* and $S \subseteq Args$. Then A defeats$_S$ B iff $(A, B) \in \mathcal{R}$ and $\neg\exists C \in S$ s.t. (C,(A, B)) $\in \mathcal{D}$.
If A defeats$_S$ B and B does not defeat$_S$ A then A *strictly* defeats$_S$ B

From hereon we may write $A \rightarrow^S B$ to denote that A defeats$_S$ B and $A \nrightarrow^S B$ to denote that A does not defeat$_S$ B.

Example 2 Let Δ be the *EAF*:

$$A \rightleftharpoons B, C \twoheadrightarrow (A \rightharpoonup B)$$

A and B defeat$_S$ each other for S = $\emptyset$, $\{A\}$ and $\{B\}$. Also, B defeats$_{\{C\}}$ A, but A does not defeat$_{\{C\}}$ B (i.e., B strictly defeats$_{\{C\}}$ A).

We state an obvious property of the defeat relation:

Proposition 1. If A defeats$_S$ B, then $\forall S' \subseteq S$, A defeats$_{S'}$ B.

We now define a conflict free set S of arguments. One might define such a set as one in which no two arguments attack each other. However, if A asymmetrically attacks B ($A \rightharpoonup B$) but an argument C claims that B is preferred to A ($C \twoheadrightarrow (A \rightharpoonup B)$) then B and A may both be justified. We may therefore want to allow *both* A and B to appear in an admissible extension of an *EAF*, which (as in the case of a Dung framework) will be defined as a conflict free set S whose contained arguments are acceptable w.r.t. S. This would clearly be inappropriate if arguments A and B refer to logically incompatible states of affairs as when arguing about what is believed to be the case. However, this may be appropriate in a practical reasoning context. Consider value based argumentation over action [8] in which an argument B justifying a course of action, such as a medical treatment, is asymmetrically attacked by an argument A claiming that the treatment is prohibitively expensive. If the value promoted by B (improving patient health) is ranked higher than that promoted by A (reducing costs), then the attack is

[2]Note that the logic programming formalisms in [34] and [24] similarly qualify the success of an attack A on B w.r.t. arguments, in some given set S, expressing priorities over rules in A and B. These works are discussed in sections 7 and 8 respectively

removed and both arguments may then be held as justified. One accepts that the action is expensive, while still pursuing the course of action (this example will be formalised in Section 5). Hence, a conflict free set is defined as follows:

Definition 6 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an *EAF*. Then $S \subseteq Args$ is conflict free iff $\forall A, B \in S$: if $(A, B) \in \mathcal{R}$, then $(B, A) \notin \mathcal{R}$ and $\exists C \in S$ s.t. $(C,(A, B)) \in \mathcal{D}$.

Intuitively, if $A, B \in S$ and A attacks B, then S is conflict free only if B does not attack A and there is a C that defence attacks the attack from A to B. Suppose the *EAF* ($Args = \{A, B, C\}$, $\mathcal{R} = \{(A, B)\}$, $\mathcal{D} = \{(C, (A, B))\}$). Then $S = \{A, B\}$ is not conflict free, but $S' = \{A, B, C\}$ is conflict free. Note that no two arguments in a conflict free set defeat each other:

Proposition 2. Let S be conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then for any $X, Y \in S$, X does not defeat$_S$ Y.

The right to left half of Proposition 2 does not hold. The following is a counter-example that will be discussed later in Section 7.

Example 3 Referring to the *EAF* in Figure 4, no two arguments in $\{S\} = \{A, B, C, D\}$ defeat$_S$ each other. However S is not conflict free since it contains arguments that symmetrically attack.

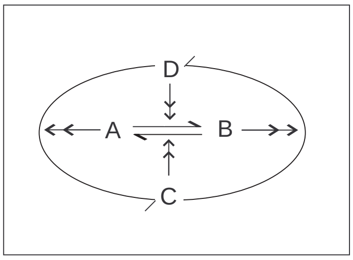


Figure 4: *EAF* for Example 3

4 Extensional Semantics for Extended Argumentation Frameworks

In this section extensional semantics for an *EAF* are defined in much the same way as for Dung frameworks. However, the definition of acceptability for *EAF*s extends Dung's definition.

4.1 Defining the Acceptability of Arguments

Consider Example 2 (see Figure 5-a). Is A acceptable w.r.t. $S = \{A\}$?. We have that $B \rightarrow^S A$, and the only argument that can reinstate A is A itself, via the defeat $A \rightarrow^S B$ denoting that A successfully attacks B w.r.t. S. However, the success of A's attack on B is under challenge by the d attack from the argument C expressing that

B is preferred to A. Hence, we need to ensure that some argument in S defeats$_S$ C, effectively 'reinstating' $A \rightarrow^S B$. One might therefore propose the following definition of acceptability:

> **Ac1:** Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an *EAF*. $A \in Args$ is acceptable w.r.t. $S \subseteq Args$ iff $\forall B \in Args$ such that $B \rightarrow^S A$, there is a $C \in S$ such that $C \rightarrow^S B$ **and** $\forall D \in Args$ s.t. (D,(C, B)) $\in \mathcal{D}$, there is a $E \in S$ s.t. $E \rightarrow^S D$.

Recall that an admissible Dung extension is a set S whose contained arguments are acceptable w.r.t. S. One can then motivate a suitably 'rational' definition of acceptability by appealing to some basic intuitions that the derived notion of admissibility should satisfy. Specifically, Dung's 'Fundamental Lemma' [19] expresses that if S is admissible, and A is acceptable w.r.t. S, then $S \cup \{A\}$ should be admissible. Intuitively, an admissible extension represents a coherent set of arguments - a 'position' - that defends (in the sense of reinstating) each of its contained arguments. If I have already established such a position S in the course of an argument with an adversary, and show that given this position I can defend another argument A that I propose, then clearly I would want that upon inclusion of A in S, A does not undermine S by making some argument in S no longer defended by S.

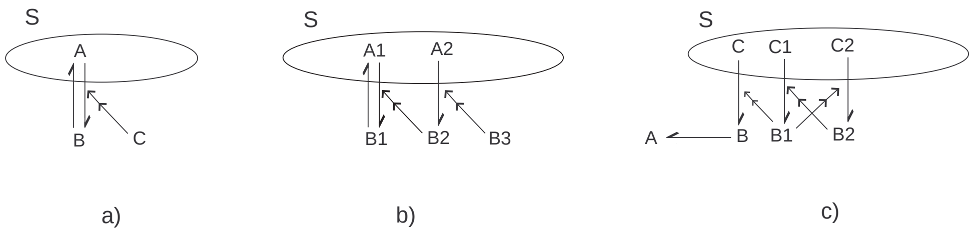


Figure 5: Examples motivating definition of acceptability for *EAF*s

Ac1's definition of acceptability, and the derived notion of admissibility, fails to satisfy the fundamental lemma. Consider Figure 5-b, in which $A1$ and $A2$ (which is not attacked by any argument) are **Ac1** acceptable w.r.t. S, and so S is **Ac1** admissible. However, $B3$ (which is not attacked by any argument) is acceptable w.r.t. S, but S' = $\{A1, A2\} \cup \{B3\}$ is not admissible since $A1$ is now not **Ac1** acceptable w.r.t. S'. This is because the inclusion of $B3$ in S' means that $A2 \nrightarrow^{S'} B2$. Dung frameworks satisfy the fundamental lemma because every attack originating from an argument in an admissible extension is preserved by inclusion of an argument acceptable w.r.t. the extension. This is not the case for *EAF*s. Inclusion of $B3$ cancels out the attack from $A2$ to $B2$, so that $A2$ no longer defeats $B2$. Hence, a definition of acceptability that ensures satisfaction of the fundamental lemma will require that when checking the acceptability of $A1$ w.r.t. S, one not only needs to check that $A1 \rightarrow^S B1$ is reinstated, but also that $A2 \rightarrow^S B2$ is reinstated. The latter is not the case since no argument in S defeats$_S$ $B3$. In general, we need to check for the existence of a defeat reinstatement *set*.

Definition 7 Let $S \subseteq Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then R_S = $\{X_1 \rightarrow^S Y_1, \ldots, X_n \rightarrow^S Y_n\}$ is a reinstatement set for $C \rightarrow^S B$, iff:

1. $C \rightarrow^S B \in R_S$
2. for $i = 1 \ldots n$, $X_i \in S$
3. $\forall X \rightarrow^S Y \in R_S$, $\forall Y'$ s.t. (Y',(X, Y)) $\in \mathcal{D}$, there is a $X' \rightarrow^S Y' \in R_S$

We now formally define the acceptability of an argument:

Definition 8 Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an *EAF*. $A \in Args$ is *acceptable* w.r.t. $S \subseteq Args$, iff:
$\forall B$ s.t. $B \rightarrow^S A$, there is a $C \in S$ s.t. $C \rightarrow^S B$ and there is a reinstatement set for $C \rightarrow^S B$.

Under this definition of acceptability, S in Figure 5-b is not admissible since $A1$ is not acceptable w.r.t. S. In Figure 5-c, A is acceptable w.r.t. S, given the reinstatement set $\{C \rightarrow^S B, C1 \rightarrow^S B1, C2 \rightarrow^S B2\}$ for $C \rightarrow^S B$. If in addition we had an argument $B3$ such that $B3 \twoheadrightarrow (C2 \rightharpoonup B2)$ and no argument in S that defeats$_S$ $B3$, then no reinstatement set for $C \rightarrow^S B$ would exist. Note that the acceptability of an argument with respect to a set S does not amount to checking whether S is admissible. In Figure 5-c, suppose an additional argument $B4$ that attacked and defeated$_S$ $C2$. A would still be acceptable w.r.t. S, but S would not be admissible.

4.2 Defining the Extensions of an Extended Argumentation Framework

The admissible, preferred, complete and stable extensions of an EAF are now defined in the same way as for Dung frameworks:

Definition 9 Let S be a conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then:

- S is an admissible extension iff every argument in S is acceptable w.r.t. S.
- S is a preferred extension iff S is a set inclusion maximal admissible extension
- S is a complete extension iff each argument which is acceptable w.r.t. S is in S
- S is a stable extension iff $\forall B \notin S$, $\exists A \in S$ such that A defeats$_S$ B

For $s \in$ {preferred, complete, stable}, A is sceptically, respectively credulously, justified under the s semantics iff A is in every, respectively at least one, s extension. Note that $\{B, C', E\}$ is the single preferred, complete, and stable extension in Example 1.

We now state some basic results that hold of Dung and Extended Argumentation Frameworks. In particular, checking for existence of a reinstatement set when defining acceptability ensures that Dung's fundamental lemma holds for *EAF*s:

Proposition 3. Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an Extended Argumentation Framework, S an admissible extension of Δ, and let A, A' be arguments which are acceptable w.r.t S. Then:

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

Proposition 4 follows immediately from Proposition 3:

Proposition 4. Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an Extended Argumentation Framework.

1. The set of all admissible extensions of Δ form a complete partial order w.r.t. set inclusion
2. For each admissible extension E of Δ there exists a preferred extension E' of Δ such that $E \subseteq E'$

Since $\emptyset$ is an admissible extension of every *EAF*, then Proposition 4 implies that:

Corollary 1. Every *EAF* possesses at least one preferred extension.

As in the case of a Dung framework, not every *EAF* has a stable extension. For example, the *EAF* $A \rightharpoonup A$ does not have a stable extension. Finally:

Proposition 5. Every stable extension of an *EAF* is a preferred extension but not vice versa.

In [19], the grounded extension of a Dung framework $AF = (Args, \mathcal{R})$ is defined as the least fixed point of the framework's characteristic function F (see Definitions 2 and 3). F is shown to be monotonic, i.e., if $S \subseteq S'$ then $F(S) \subseteq F(S')$, where S and S' are both subsets of $Args$. The monotonicity of F guarantees the existence of a least fixed point and gives it a constructive flavour: its least fixed point - the AF's grounded extension ($GE(AF)$) - can be approached under iterated application of F to the empty set. Defining the sequence:

- $F^0 = \emptyset$
- $F^{i+1} = F(F^i)$

then $GE(AF) \subseteq \bigcup_{i=0}^{\infty}(F^i)$.

If AF is *finitary*, i.e., every argument is attacked by at most a finite number of arguments, then the grounded extension can be obtained by iterative application to the empty set:

$GE(AF) = \bigcup_{i=0}^{\infty}(F^i)$

We now define the characteristic function of an *EAF*. For reasons that will become apparent later, the domain of the function is restricted to conflict free sets.

Definition 10 Let $\Delta = (Args, \mathcal{R}, \mathcal{D})$ be an *EAF*, $S \subseteq Args$, and 2^{ArgsC} denote the set of all conflict free subsets of Args. The characteristic function F_Δ of Δ is defined as follows:

- $F_\Delta : 2^{ArgsC} \mapsto 2^{Args}$
- $F_\Delta(S) = \{A | A \text{ is acceptable w.r.t } S\}$

From hereon, since we will always refer to an arbitrary but fixed *EAF*, we will simply write F rather than F_Δ. As for Dung frameworks, for any conflict free $S \subseteq Args$ in $\Delta = (Args, \mathcal{R}, \mathcal{D})$, S is admissible iff $S \subseteq F(S)$, and complete iff S is a fixed point of F. We now define a procedure iteratively applying F to an *EAF*:

Definition 11 Define for any *EAF* $(Args, \mathcal{R}, \mathcal{D})$ the following sequence of subsets of $Args$.

- $F^0 = \emptyset$
- $F^{i+1} = F(F^i)$

However, an *EAF*s characteristic function F is **not**, in general, monotonic. For example, A is acceptable w.r.t. $S = \{C, C1, C2\}$ in Figure 5-c), but is not acceptable w.r.t. the conflict free $\{C, C1, C2, B1, B2\}$.

Since in general F is not monotonic, existence of a least fixed point is not guaranteed. However Proposition 6 states that each F^i in Definition 11 is conflict free (hence F can be applied iteratively), and that the sequence is monotonically increasing under $\subseteq$.

Proposition 6 Let F be the characteristic function of an *EAF*, and $F^0 = \emptyset$, $F^{i+1} = F(F^i)$. Then $\forall i$, $F^i \subseteq F^{i+1}$ and F^i is conflict free.

Hence, rather than define the grounded extension as the least fixed point, the grounded extension of a *finitary EAF* is defined in terms of the sequence in Definition 11.

Definition 12 $(Args, \mathcal{R}, \mathcal{D})$ is finitary iff $\forall A \in Args$, the set $\{B|(B, A) \in \mathcal{R}\}$ is finite, and $\forall (A, B) \in \mathcal{R}$, the set $\{C|(C, (A, B)) \in \mathcal{D}\}$ is finite.

Definition 13 Let Δ be a finitary *EAF* and $F^0 = \emptyset$, $F^{i+1} = F(F^i)$. Then $\bigcup_{i=0}^{\infty}(F^i)$ is the grounded extension $GE(\Delta)$ of Δ

It should be obvious to see that the grounded extension of any *EAF* $(Args, \mathcal{R}, \mathcal{D})$ will contain any argument that is not attacked (i.e. the set $\{X \mid \neg\exists Y, (Y, X) \in \mathcal{R}\}$). The following example illustrates that in general it does not hold that the grounded extension is a subset of the sceptically justified arguments under the preferred semantics (a result that does hold for Dung frameworks).

Example 4 $C \rightharpoonup B \rightharpoonup A$, $B \twoheadrightarrow (C \rightharpoonup B)$.

The 'self reinstating argument' B is excluded when adopting a 'constructivist' approach: $F^1 = \{C\}$, $F^2 = \{C, A\}$, $F^3 = \{C, A\}$ is the grounded extension. However, adopting a 'declarative' perspective, $\{C\}$, $\{C, A\}$ and $\{C, B\}$ are admissible, and $\{C, A\}$ and $\{C, B\}$ the preferred extensions. Hence, only C is sceptically justified under the preferred semantics. The grounded semantics excludes self-reinstating arguments such as B, and so gives justified status to arguments such as A.

To conclude, we have extended Dung's theory of argumentation to accommodate metalevel arguments that express preferences between arguments. The extended theory preserves the abstract nature of Dung's approach, and based on a novel notion of acceptability that additionally requires reinstatement of defeats, defines the extensional semantics in the same way (apart from the grounded semantics) as for a Dung framework. We have also shown that some of the basic results that hold of the extensional semantics for Dung frameworks, also hold for the extended frameworks. We therefore suggest that applications and developments of Dung's work can be readily modified to accommodate argumentation about preferences. For example, game based proof theories (e.g., [15, 37]) and dialogue models that build on these games (e.g., [33]), make core use of the notion of reinstatement encoded in the acceptability calculus. Notice also, that Proposition 4 implies that (as with Dung's theory) when defining games for determining the credulous acceptance of an argument, it will suffice to show that A belongs to an admissible extension rather than having to construct a preferred extension containing A. Finally, note that while in general the characteristic function of an *EAF* is not monotonic, the following two sections study two special classes of *EAF* whose characteristic functions are monotonic.

5 Value Based Argumentation in Hierarchical Extended Argumentation Frameworks

In this section we study a special class of *EAF* - *hierarchical EAF*s - and then show how Bench-Capon's value based argumentation [11] can be formalised in *hierarchical EAF*s, and extended to integrate metalevel reasoning about values and value orderings.

Thus far we have considered *EAF*s in which no restrictions are placed on the interaction between object level arguments and metalevel arguments expressing preferences over the object level arguments. In hierarchical *EAF*s, the argumentation is 'stratified' into levels so that, intuitively, each level is a Dung framework in which all binary attacks are restricted to arguments within the framework. These binary attacks are then attacked by defence attacks that exclusively originate from arguments expressing preferences in the immediate metalevel.

Definition 14 $\Delta = (Args, \mathcal{R}, \mathcal{D})$ is a *hierarchical EAF* iff there exists a partition Δ_H = ($((Args_1, \mathcal{R}_1), \mathcal{D}_1)$, ..., $((Args_j, \mathcal{R}_j), \mathcal{D}_j)$, ...) such that:

- $Args = \bigcup_{i=1}^{\infty} Args_i$, $\mathcal{R} = \bigcup_{i=1}^{\infty} \mathcal{R}_i$, $\mathcal{D} = \bigcup_{i=1}^{\infty} \mathcal{D}_i$, and for $i = 1 \ldots \infty$, $(Args_i, \mathcal{R}_i)$ is a Dung argumentation framework.
- $(C,(A,B)) \in \mathcal{D}_i$ implies $(A,B) \in \mathcal{R}_i$, $C \in Args_{i+1}$

Δ is a *bounded hierarchical EAF* iff its partition Δ_H is of the form ($((Args_1, \mathcal{R}_1), \mathcal{D}_1)$, ..., $((Args_n, \mathcal{R}_n), \mathcal{D}_n)$), where $\mathcal{D}_n = \emptyset$

The characteristic function of a bounded hierarchical *EAF* satisfies monotonicity:

Proposition 7. Let F be the characteristic function of a bounded hierarchical *EAF* $(Args, \mathcal{R}, \mathcal{D})$. Let S and S' be conflict free subsets of $Args$ such that $S \subseteq S'$. Then $F(S) \subseteq F(S')$.

If F's domain were not restricted to conflict free sets then F would not be monotonic. Consider the following counter-example that assumes an unrestricted domain:

Example 5 Let $\Delta = C \rightleftharpoons B \rightharpoonup A$, $E \rightharpoonup D$, $D \twoheadrightarrow (C \rightharpoonup B)$

$A \in F(\{C, E\})$ since although $B \rightarrow^{\{C,E\}} A$, we have $C \rightarrow^{\{C,E\}} B$ for which there is a reinstatement set $\{C \rightarrow^{\{C,E\}} B, E \rightarrow^{\{C,E\}} D\}$. However, $A \notin F(\{C, E, D\})$ where $\{C, E, D\}$ is not conflict free. $B \rightarrow^{\{C,E,D\}} A$ but $C \nrightarrow^{\{C,E,D\}} B$.

The monotonicity of F guarantees existence of a least fixed point.

Definition 15 The grounded extension of a bounded hierarchical *EAF* Δ, denoted $GE(\Delta)$, is the least fixed point of its characteristic function F.

Hence, as in the case of a Dung framework, the grounded extension can be approached, and under the finitary restriction be obtained, by iterative application to the empty set. Note that Proposition 6 implies that the iteration defined in Definition 11 is indeed possible, i.e., that each F^i is conflict free.

Proposition 8 Let Δ = $(Args, \mathcal{R}, \mathcal{D})$ be a bounded hierarchical *EAF*, $F^0 = \emptyset$, $F^{i+1} = F(F^i)$. Then:

1. $\bigcup_{i=0}^{\infty}(F^i) \subseteq GE(\Delta)$

2. If Δ is finitary, then $\bigcup_{i=0}^{\infty}(F^i) = GE(\Delta)$

The extended argumentation theory provides a unifying framework in which to formalise and extend works that augment Dung's framework with preferences [3, 11]. In support of this claim we formalise and extend value based argumentation in hierarchical *EAF*s. Firstly, we recall [11]'s value based argumentation:

Definition 16 $\langle Args,\mathcal{R},V,val,P\rangle$ is a value-based argumentation framework (VAF), where val is a function from $Args$ to a non-empty set of values V, and P is a set $\{a_1,\ldots,a_n\}$, where each a_i names a total ordering (audience) $>_{a_i}$ on $V \times V$.
An *audience specific* VAF ($aVAF$) is a 5-tuple $\langle Args,\mathcal{R},V,val,a\rangle$ where $a \in P$.

Definition 17 Let $\langle Args,\mathcal{R},V,val,a\rangle$ be an $aVAF$. Then $A \in Args$ defeats$_a$ $B \in Args$ iff $(A, B) \in \mathcal{R}$ and it is not the case that $val(B) >_a val(A)$.
We say that $(A, B) \in \mathcal{R}_a$ iff A defeats$_a$ B.

The extensions and justified arguments of an $aVAF$ $\Gamma = \langle Args,\mathcal{R},V,val,a\rangle$ are those of the Dung framework $(Args, \mathcal{R}_a)$. Symmetric attacks in Γ are thus resolved to obtain asymmetric defeats in $(Args, \mathcal{R}_a)$, and (as discussed in Section 3) asymmetric attacks may be removed so that if A asymmetrically attacks B, but B's value is ordered (by audience $>_a$) above A's value, then neither argument defeats$_a$ each other, and so both may appear in the same extension. [11] shows that if for every $(A, B) \in \mathcal{R}$ either $val(A) >_a val(B)$ or $val(B) >_a val(A)$, and assuming no cycles in the same value in Γ, then there is guaranteed to be a unique, non-empty preferred extension of Γ, and a polynomial time algorithm to find it. Thus, for problematic odd cycles such as $A \rightharpoonup B \rightharpoonup C \rightharpoonup A$ which only yield $\emptyset$ as an admissible extension of a Dung framework, then if A, B and C do not promote the same value, any audience derived defeat relation will yield a single preferred extension.

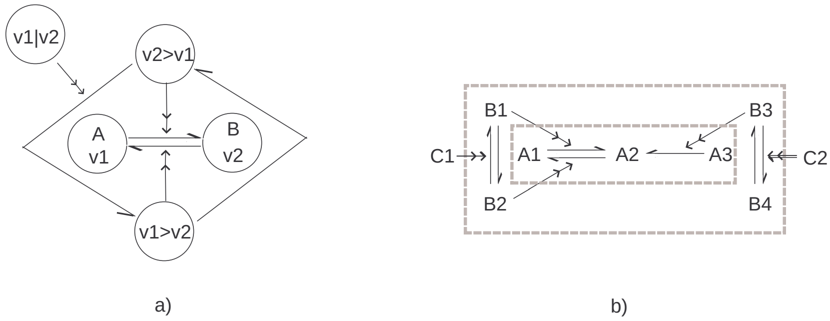


Figure 6: Integrating meta-level argumentation about values and value orderings

Reasoning *about* values and value orderings can be accommodated by formalising value based argumentation in a hierarchical *EAF* in which the first level includes the arguments and attacks in an $aVAF$. Pairwise orderings on values in V are *value preference arguments* in the second level, so that if $v1, v2 \in V$ then '$v1 > v2$' and '$v2 > v1$' are value preference arguments that attack each other in the second level (see Figure

6a)). If level 1 arguments A and B respectively promote $v1$ and $v2$, and $B \rightharpoonup A$, then $(v1 > v2) \twoheadrightarrow (B \rightharpoonup A)$. Finally, an audience is represented by a level 3 *audience argument* that denotes a choice of ordering, and thus defence attacks the binary attacks between value preference arguments. Thus, $v1|v2 \twoheadrightarrow ((v2 > v1) \rightharpoonup (v1 > v2))$, where $v1|v2$ is a level 3 argument representing an audience that selects $v1 > v2$. Now, the unique preferred extension of the *EAF* in Figure 6a) is $\{A, v1 > v2, v1|v2\}$. In this way, we can represent the meta-level reasoning required to find the preferred extension of an $aVAF$.

Definition 18 Let Γ be an $aVAF$ $\langle Args,\mathcal{R},V,val,a\rangle$. Then the EAF

$$\Delta = (Args1 \cup Args2 \cup Args3, \mathcal{R}1 \cup \mathcal{R}2 \cup \mathcal{R}3, \mathcal{D}1 \cup \mathcal{D}2 \cup \mathcal{D}3)$$

is defined as follows:

1. $Args1 = Args$, $\mathcal{R}1 = \mathcal{R}$
2. $\{v > v'|v, v' \in V, v \neq v'\} \subseteq Args2$
   $\{(v > v', v' > v) \mid v > v', v' > v \in Args2\} \subseteq \mathcal{R}2$
3. $\{a\} \subseteq Args3$, $\emptyset \subseteq \mathcal{R}3$
4. - $\{ (v > v',(A, B)) \mid (A, B) \in \mathcal{R}1$, $val(B) = v$, $val(A) = v'\} \subseteq \mathcal{D}1$
     (i.e., value preference arguments in $Args2$ defence attack attacks between arguments in $Args1$)
   - $\{ (a, (v > v', v' > v)) \mid a \in Args3, (v > v', v' > v) \in \mathcal{R}2, v' >_a v\} \subseteq \mathcal{D}2$
     (i.e., audience arguments in $Args3$ defence attack attacks between value preference arguments in $Args2$)
   - $\mathcal{D}3 = \emptyset$ (the framework is bounded)

If in 2,3 and 4, the $\subseteq$ relation is replaced by $=$, then Γ and Δ are said to be *equivalent*.

Proposition 9. Let $\Gamma = \langle Args,\mathcal{R},V,val,a\rangle$ be an $aVAF$ and Δ its equivalent EAF. Let $s \in$ {preferred, grounded, complete}. Then $\forall A \in Args$, A is a sceptically, respectively credulously, justified argument of Γ under the s semantics, iff A is a sceptically, respectively credulously, justified argument of Δ under the s semantics.

Notice that Δ is defined so that one could additionally consider other arguments and attacks in levels 2 and 3. For example, arguments in level 2 that directly attack value preference arguments, or arguments in level 3 representing different audiences.

Indeed, given a VAF $\langle Args,\mathcal{R},V,val,P\rangle$, then its *EAF* is obtained as in Definition 18, except that now $\{a|a \in P\} \subseteq Args3$ (recall that P is the set of all possible audiences). A *VAF* and its obtained *EAF* are then said to be *equivalent* if $\{a|a \in P\} = Args3$. Notice that if for any $a, a' \in P$, $(a, (v > v', v' > v))$, $(a', (v' > v, v > v'))$ $\in \mathcal{D}2$, then it follows from the definition of an *EAF* (Definition 4) that a and a' attack each other, i.e. (a, a'), $(a', a) \in \mathcal{R}3$. Since each possible audience argument corresponds to a different total ordering on values, and $\forall v, v' \in V$, $v > v'$ and $v' > v$ are value preference arguments in $Args2$, then every audience argument will $\mathcal{R}3$ attack every other audience argument. Intuitively, a level 1 argument in the *EAF* will then be sceptically justified under the preferred semantics iff it is *objectively accepted* [11] (justified irrespective of the chosen audience) in the corresponding *VAF*. A level 1 argument will be credulously justified iff it is *subjectively accepted* [11] (justified for at least one chosen audience). Of course, given representation of multiple audiences, one can in principle incorporate a fourth level to model argumentation over preferences between audiences [3]. Notice, that an audience argument is not necessarily restricted to instantiation by the value ordering that constitutes the audience, but may be instantiated by some justification for the ordering that is the claim of the argument. Furthermore, Definition 18 also allows for arguments that express preferences that are not derived on the basis of value orderings. We now illustrate the above ideas with an example of value based argumentation over action.

Example 6 In what follows, we assume instantiation of level 1 arguments about actions ($A1, A2, A3$ in Figure 6b)) in a BDI logic, as described in [8]. $A1$ and $A2$ are arguments for the medical actions 'give aspirin' and 'give chlopidogrel' respectively. These arguments relate the current beliefs that warrant (are preconditions for) the actions bringing about states of affairs that realise a desired goal and so appeal to a value. Both $A1$ and $A2$ appeal to the value of *improving the patient's health*. They symmetrically attack since they claim alternative actions for the goal of preventing blood clotting.

In [27] we describe construction of level 2 arguments that claim preferences between level 1 arguments based on the relative *degree of promotion* of a desired value by the level 1 arguments. Thus, $B1$ is an argument based on clinical trial 1's conclusion that $A2$'s chlopidogrel is more efficacious than $A1$'s aspirin at preventing blood clotting. Hence $B1 \twoheadrightarrow (A1 \rightharpoonup A2)$. Suppose also $B2$ based on clinical trial 2's conclusion that the opposite is the case. Hence $B1 \rightleftharpoons B2$. At this stage neither $A1$ or $A2$ are sceptically justified under any semantics. However $C1$ claims that trial 1 is preferred to trial 2 since the former uses a more statistically robust design. Now $A2$ and not $A1$ is sceptically justified.

However, $A3$ appealing to the value of *cost*, states that chlopidogrel is prohibitively expensive and so asymmetrically attacks $A2$. $B3$ is the value preference argument ordering *improving the patient's health* over *cost*, and $B4$ the value preference argument ordering *cost* over *improving the patient's health*. Now, neither $A2$ or $A1$ are sceptically justified (notice that now there is a preferred extension containing $B3$, $A2$ *and* $A3$). Finally, $C2$ is an audience argument that selects $B4$'s value preference over $B3$. This choice is justified by appealing to the utilitarian principle: since financial resources are low, and chlopidogrel is costly, then use of chlopidogrel will compromise treatment of other patients, and so one should preferentially order *cost* over *improving the patient's health* (such a trade off is often made in medical contexts). Hence, $A3$'s attack on $A2$ succeeds, and so $A1$ is now sceptically justified; aspirin is now the preferred course of action.

In this section we have integrated metalevel reasoning about values and value orderings in a hierarchical *EAF*. The question naturally arises as to whether it is always possible, or indeed desirable, to stratify argumentation according to the hierarchical prescription. Pragmatically, it may well be computationally expensive to ensure that for a given knowledge base, the arguments and attacks defined conform to the hierarchical restriction. Philosophically, a bounded hierarchical model assumes a highest level of reasoning in which knowledge is certain and so immune to doubt and conflict. For example, for value based argumentation, a three level hierarchical framework assumes that preferences over value orderings (audiences) are given, and not themselves

[3] One could in this way express constraints on value preferences that effectively identify classes of audiences; for example an argument preferring audiences that order religious over secular values will denote the religious class of audiences.

subject to reasoning. To avoid such an assumption, while avoiding having to ascend levels to infinity, one can appeal to the intuition that our metalevel reasoning is often informed by our object level reasoning; argumentation at lower levels may thus impact on argumentation at higher levels. As Searle puts it [36], values are often the product of practical reasoning rather than an input to it. Indeed, we refer the reader to an intuitive illustration of non-hierarchical extended argumentation in Example 14 (Section 7).

6 Preference Symmetric Extended Argumentation Frameworks

This section focuses on *preference symmetric EAF*s in which only attacks between symmetrically attacking arguments can be attacked by arguments expressing preferences. This restriction is satisfied by formalisms in which all attacks between arguments are symmetric (e.g. [16]), and formalisms in which asymmetric attacks succeed as defeats irrespective of preferences between arguments. An example of the latter is the argument based logic programming formalism in [34], in which A asymmetrically attacks B iff A claims (proves) what was assumed non-provable (through negation as failure) by a rule in B. This attack succeeds as a defeat irrespective of whether B is preferred to A. The formalism of [34] will be described and formalised as an instance of a *preference symmetric EAF* in Section 7.

Definition 19 $\Delta = (Args, \mathcal{R}, \mathcal{D})$ is a *preference symmetric EAF* iff: if $(C, (B, A)) \in \mathcal{D}$, then $(A, B) \in \mathcal{R}$.

No conflict free set and so admissible extension S of a *preference symmetric EAF* (*ps-EAF*) Δ can contain arguments that attack each other. Suppose such a set. Then by Definition 6, if for $A, B \in S$, $(B, A) \in \mathcal{R}$, then $\exists C \in S$ such that $(C, (B, A)) \in \mathcal{D}$, and it must be the case that $(A, B) \notin \mathcal{R}$, contradicting Δ is a *ps-EAF*.

As already noted, value based argumentation does not conform to the restriction imposed by *ps-EAF*s. Hence, it may be that $(B, A) \in \mathcal{R}$ and yet B and A both appear in an admissible extension and are both justified. As discussed in Section 3 and illustrated in Example 6, this may be appropriate when applying argumentation to practical reasoning, where an argument for an action may be accepted while acknowledging the validity of a challenge represented by an asymmetrically attacking argument. However, it is not appropriate when the arguments relate to inherently contradictory states of affairs. Consider Pollock's argumentation system [32] in which if B 'Pollock undercuts' A, and A is preferred to B, then B does not defeat A. The standard example describes A to be an argument claiming that an object is red given that it looks red, and the rule that things that look red are red. This argument is undercut (the application of the rule is invalidated) by an argument B claiming that the object is illuminated by red light. Now, if A is preferred to B, then A and B may both be justified. This is somewhat strange; that one should accept together an argument that makes use of a rule, and an argument that invalidates use of the rule [4]. We suggest that the Pollock undercut be reformulated, either as a symmetric attack, or as an asymmetric attack that succeeds as a defeat irrespective of the relative preference.

Note that we do not claim that the *only* way to ensure that attacking arguments do not appear in the same admissible extension is by requiring that the arguments and

[4] Notice that in the usual formalisation of Pollock undercuts, A's rule is represented as a defeasible implication $\alpha \Rightarrow \beta$ (where α and β are wff in some first order language) and B's claim is of the form $\neg(\alpha \Rightarrow \beta)$

attacks instantiate a *ps-EAF*. It may well be that this property is satisfied given the way in which arguments, attacks and preferences are defined in the underlying logic. For example, preference based argumentation frameworks [3] are instantiated by arguments defined on the basis of a propositional knowledge base. Only asymmetric attacks are defined, and a preference relation on arguments may result in the removal of these attacks. However, in [2] it is shown that if the preference relation applied satisfies some intuitively desirable properties, then no two arguments with contradictory propositions will be jointly justified.

We now present some results that hold for *preference symmetric EAF*s. Firstly, recall our motivation of the definition of acceptability for *EAF*s in Section 4.1. We proposed a definition of acceptability (**Ac1**) that is more formally defined here as 'preference symmetric', or *ps*, acceptability.

Definition 20 Let $(Args, \mathcal{R}, \mathcal{D})$ be a *EAF*. Then for any $S \subseteq Args$:

1. $C \rightarrow^S B$ is *reinstated* iff $\forall D \in Args$ s.t. $(D,(C,B)) \in \mathcal{D}$ there is a $E \in S$ s.t. $E \rightarrow^S D$.
2. $A \in Args$ is *ps* acceptable w.r.t. S iff $\forall B \in Args$ such that $B \rightarrow^S A$, there is a $C \in S$ such that $C \rightarrow^S B$, and $C \rightarrow^S B$ is reinstated.

We saw that when evaluating the acceptability of $A1$ w.r.t. S in Figure 5b), *ps* acceptability does not suffice. We are required to not only check that $A1 \rightarrow^S B1$ is reinstated, but also $A2 \rightarrow^S B2$, since otherwise S would be admissible, and inclusion of the acceptable $B3$ in S would result in an S' that is not admissible. However, the requirement that one checks for reinstatement of $A2 \rightarrow^S B2$ would not arise if $B2 \rightharpoonup A2$, as would be the case in a *ps-EAF*. This is because S would not be admissible, since $A2$ would not be acceptable w.r.t S (under *ps* acceptability or the more general definition). Indeed, *ps* acceptability and the general definition of acceptability in Definition 8 coincide for *ps-EAF*s, in the sense that:

Proposition 10. Let $(Args, \mathcal{R}, \mathcal{D})$ be a preference symmetric *EAF*. Let S be a conflict free subset of $Args$. Then, every argument in S is *ps* acceptable w.r.t. S iff every argument in S is acceptable w.r.t. S.

Proposition 10 implies that extensional semantics for *ps-EAF*s can equivalently be defined on the basis of ps acceptability. Furthermore, the characteristic function of a *ps-EAF* is monotonic. We can therefore define the grounded extension of a *ps-EAF* in terms of its characteristic function's least fixed point, and under the finitary restriction obtain the grounded extension by iterative application to the empty set.

Proposition 11. Let F be the characteristic function of a preference symmetric *EAF* $(Args, \mathcal{R}, \mathcal{D})$. Let S and S' be conflict free subsets of $Args$ such that $S \subseteq S'$. Then $F(S) \subseteq F(S')$.

Definition 21 The grounded extension of a preference symmetric *EAF* Δ, denoted $GE(\Delta)$, is the least fixed point of its characteristic function F.

Proposition 12 Let $\Delta = (Args, \mathcal{R}, \mathcal{D})$ be a preference symmetric *EAF*, $F^0 = \emptyset$, $F^{i+1} = F(F^i)$. Then:

1. $\bigcup_{i=0}^{\infty}(F^i) \subseteq GE(\Delta)$

2. If Δ is finitary, then $\bigcup_{i=0}^{\infty}(F^i)$ = $GE(\Delta)$

Example 7 For Δ in Figure 7, F^1 = $\{A, H\}$, F^2 = $\{A, H, G\}$, F^3 = $\{A, H, G, C\}$, F^4 = $\{A, H, G, C, E\}$ where F(F^4) = F^4 is the grounded extension.

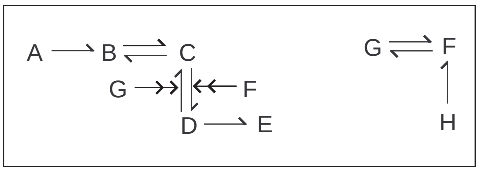


Figure 7: $\{A, H, G, C, E\}$ is the grounded extension of the *ps-EAF*

We now show that the grounded extension of a *ps-EAF* can be obtained on the basis of 'strict acceptability' (recall the definition of strict defeat$_S$ given in Definition 5):

Definition 22 Let $S \subseteq Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then A is strict-acceptable w.r.t. S iff $\forall B$ s.t. B defeats$_S$ A, $\exists$ $C \in S$ s.t. C strictly defeats$_S$ B.

Definition 23 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$), $S \subseteq Args$, and 2^{ArgsC} denote the set of all conflict free subsets of $Args$. The characteristic function F_{st} of Δ is defined as:

- F_{st} : $2^{ArgsC} \mapsto 2^{Args}$
- $F_{st}(S)$ = $\{A | A$ is strictly acceptable w.r.t $S\}$

Define for any *EAF* the following sequence of subsets of $Args$.

- $F{*}^0$ = $\emptyset$
- $F{*}^{i+1}$ = $F_{st}(F{*}^i)$

Proposition 13 states that for a *ps-EAF*, the iteration obtained by the characteristic function F based on the standard definition of acceptability, is the same as the iteration obtained by F_{st} based on strict acceptability. This implies that the grounded extension of a finitary *ps-EAF* is given by $\bigcup_{i=0}^{\infty}(F{*}^i)$.

Proposition 13 Let Δ be a preference symmetric *EAF*, and let F^i be defined as in Definition 11, $F{*}^i$ defined as in Definition 23. Then $F{*}^i$ = F^i.

Referring to the *ps-EAF* in Example 7, one can see that each argument in F^{i+1} - F^i is strictly acceptable w.r.t. F^i. For example, E is strict-acceptable w.r.t. F^3: D defeats$_{F^3}$ E, and since $G \in$ F^3, then D does not defeat$_{F^3}$ C, and so C strictly defeats$_{F^3}$ D.

7 Formalising Logic Programming with Defeasible Priorities in an Extended Argumentation Framework

Extended argumentation provides an abstract characterisation of preference application while retaining the basic conceptual machinery of Dung. We therefore propose that it provides a semantics for object level non-monotonic formalisms that accommodate defeasible reasoning about preferences (e.g. [14, 24, 34]). Arguments and attacks defined by these formalisms can instantiate an *EAF* and the justified arguments under all of Dung's semantics can be evaluated. In this section we show how the inferences obtained in Prakken and Sartor's argument based logic programming with defeasible priorities (ALP-DP) [34], equate with the claims of justified arguments of a preference symmetric *EAF* instantiated by arguments and attacks defined by an ALP-DP theory. Furthermore, unlike [34], one can then determine the justified arguments of an ALP-DP theory under **all** of Dung's extensional semantics.

7.1 Defining Arguments on the basis of an ALP-DP theory

In ALP-DP, arguments are chained sequences of named logic program rules. The head of a rule can express a priority ordering on the names of other rules. Hence, given arguments $A1$ and $A2$, one can identify 'priority' arguments $B1, B2, \ldots$ such that an ordering on rules in $A1$ and $A2$ is defined by the heads of rules in $B1, B2, \ldots$. One can then define a preference relation $\leq$ on $A1$ and $A2$ based on the ordering claimed by $B1, B2, \ldots$. In ALP-DP, rules are defeasible or strict, where only the former express debatable information.

Definition 24 An ALP-DP theory is a tuple (S, D), where:
- S is a set of strict rules of the form $s : L_0 \wedge \ldots \wedge L_m \rightarrow L_n$;
- D is a set of defeasible rules $r : L_0 \wedge \ldots \wedge L_j \wedge \sim L_k \wedge \ldots \wedge \sim L_m \Rightarrow L_n$, and:

- Each rule name r (s) is a first order term.
- Each L_i is a strong literal, i.e., an atomic first order formula, or such a formula preceded by strong negation $\neg$
- Each $\sim L_i$ is a weak literal, where L_i is a strong literal, and $\sim$ denotes negation as failure, so that $\sim L_i$ is read as "there is no evidence that L_i is the case"

As usual, a rule with variables is a scheme standing for all its ground instances. Henceforth, *head*(r) denotes the consequent L_n of the rule named r. Also, for any atom A, we say that A and $\neg A$ are the complement of each other. In the metalanguage, $\overline{L}$ denotes the complement of a literal L. The language is assumed to contain a two-place predicate symbol $\prec$ for expressing priorities on rule names. To ensure that $\prec$ is a strict partial order, the strict rules S are always assumed to contain the following:

- $o1 : (x \prec y) \wedge (y \prec z) \rightarrow (x \prec z)$
- $o2 : (x \prec y) \wedge \neg(x \prec z) \rightarrow \neg(y \prec z)$
- $o3 : (y \prec z) \wedge \neg(x \prec z) \rightarrow \neg(x \prec y)$
- $o4 : (x \prec y) \rightarrow \neg(y \prec x)$

Definition 25 An argument A based on the theory (S, D) is:

1. a finite sequence $s = [r_0, \ldots, r_n]$ of ground instances of rules such that:

   (a) for every i $(0 \leq i \leq n)$, for every strong literal L in the antecedent of r_i there is a $k < i$ such that *head*$(r_k) = L$.
   If *head*$(r_n) = x \prec y$ then A is called a 'singleton priority argument'

   (b) no two distinct rules in the sequence have the same head

   or:

2. a finite sequence $[s_1, \ldots, s_n]$, such that for $i = 1 \ldots n$, s_i is a singleton priority argument. A is said to be a 'composite priority argument' that concludes the ordering $\bigcup_{i=1}^{n}$ *head*(r_i), where $r_1, \ldots, r_n$ are, respectively, the last rules in the singleton priority arguments $s_1, \ldots, s_n$.

An argument A is based on the theory (S, D) iff all rules in A are in $S \cup D$. In the definitions that follow we assume that arguments are relative to a theory (S, D).

Definition 25-1a says that arguments formed by chaining rules can ignore weak literals. Definition 25-1b prevents arguments from containing circular chains of reasoning. In [34], arguments are exclusively defined by 25-1. Preferences between arguments are then parameterised w.r.t. a set T, based on the ordering claimed by the *set* of singleton priority arguments in T. In this paper, we have additionally defined composite priority arguments in 25-2, so that an ordering, and hence a preference, is claimed by a single argument. This is the only modification we introduce to ALP-DP as defined in [34]. From hereon, to enhance readability we will describe only propositional examples, and as an abuse of notation write arguments as sequences of rule names rather than the rules that the names identify.

Example 8 Let $S = \{o1, \ldots, o4\}$ and D the set of rules:

r1 : $\Rightarrow a$, r2 : $\Rightarrow \neg a$, r3 : $a \Rightarrow b$, r4 : $\neg a \Rightarrow \neg b$,
r5 : $\Rightarrow r2 \prec r1$, r6 : $\Rightarrow r1 \prec r2$, r7 : $\Rightarrow r4 \prec r3$, r8 : $\Rightarrow r6 \prec r5$

Amongst the arguments that can be constructed are:

$A1 = [r1], A2 = [r2], A3 = [r1, r3], A4 = [r2, r4], B1 = [r5], B2 = [r6], B3 = [r5, r7], B4 = [r6, r7], B5 = [r7], C1 = [r8]$

Note that $B3$ is a composite argument concluding the ordering $r2 \prec r1, r4 \prec r3$

Example 9 Let $S = \{o1, \ldots, o4\}$ and D the set of rules:

r1 : $\sim b \Rightarrow a$, r2 : $\Rightarrow b$, r3 : $b \Rightarrow \neg a$, r4 : $\Rightarrow r3 \prec r1$

Amongst the arguments that can be constructed are:

$A1 = [r1], A2 = [r2], A3 = [r2, r3], B1 = [r4]$.

In [34] the following is defined:

Definition 26 For any arguments A, A' and literal L:

- A is strict iff it does not contain any defeasible rule; it is defeasible otherwise.
- A' is a sub-argument of A iff A' is a subsequence of A.
- L is a conclusion of A iff L is the head of some rule in A
- L is an assumption of A iff $\sim \overline{L}$ occurs in some antecedent of a rule in A.
- If T is a sequence of rules, then $A + T$ is the concatenation of A and T
- By definition, A = *[ ]* is an argument of any ALP-DP theory, and is referred to as the *empty* argument.

7.2 Defining Attacks and Preferences Between ALP-DP Arguments

In [34], the authors motivate definition of attacks between arguments that account for the ways in which arguments can be extended with strict rules:

Definition 27 $A1$ attacks $A2$ iff there are sequences S1 and S2 of strict rules such that $A1 + S1$ is an argument with conclusion L and

1. $A2 + S2$ is an argument with a conclusion $\overline{L}$, in which case $A1$ and $A2$ are said to symmetrically *conclusion-conclusion* attack on the the pair (L,$\overline{L}$); or

2. $A2$ is an argument with an assumption $\overline{L}$, in which case $A1$ is said to *undercut* $A2$ on the pair (L,$\overline{L}$)

Example 10 In Example 8, $B3$ = [r5, r7] and $B4$ = [r6, r7] conclusion-conclusion attack since [r5, r7] has the conclusion $r2 \prec r1$ and [r6, r7] + [$o4$] has the conclusion $\neg(r2 \prec r1)$. Also, $A3$ and $A4$ conclusion-conclusion attack on the pairs $(a, \neg a)$ and $(b, \neg b)$. In Example 9, $A1$ and $A3$ symmetrically conclusion-conclusion attack on the pair $(a, \neg a)$, and both $A2$ and $A3$ undercut $A1$ on the pair (b,$\neg b$).

In general, since arguments A and A' can attack each other on more than one pair of conclusions, then there may different sequences of strict rules that extend these arguments to account for the pairs of conclusions on which they attack. Note that arguments can conclusion-conclusion attack or undercut themselves. For example, $[r :\sim a \Rightarrow a]$ undercuts itself, and $[r1 :\Rightarrow a, r2 : a \Rightarrow \neg a]$ conclusion-conclusion attacks itself.

Notation 2 A is *incoherent* iff A conclusion-conclusion attacks or undercuts itself.

In [34], a preference amongst conclusion-conclusion attacking arguments is based on a comparison of the *sets* of defeasible rules that contribute to derivation of the conflicting conclusions.

Definition 28 If $A+S$ is an argument with conclusion L, the defeasible rules $R_L(A+S)$ **relevant** to L are:

1. $\{r_d\}$ iff A includes defeasible rule r_d with head L

2. $R_{L_1}(A+S) \cup \ldots \cup R_{L_n}(A+S)$ iff A is defeasible and S includes a strict rule $s : L_1 \wedge \ldots \wedge L_n \rightarrow L$

We define [34]'s ordering on the above sets, and hence preferences amongst arguments,w.r.t. an ordering concluded by a single composite priority argument (rather than a set of singleton priority arguments as in [34]):

Definition 29 Let C be a priority argument concluding the ordering $\prec$. Let R and R' be sets of defeasible rules. Then $R' > R$ iff $\exists r \in R$ s.t. $\forall r'$, $r' \in R'$ implies $r \prec r'$.

Intuitively, R can be made better by replacing some rule in R with any rule in R', while the reverse is impossible. Given a priority ordering $\prec$ concluded by an argument C, we say that A is preferred$_\prec$ to B if for every pair (L, L') on which they conclusion-conclusion attack, the set of A's defeasible rules relevant to L is stronger ($>$) than the set of B's defeasible rules relevant to L'.

Definition 30 Let C be a priority argument concluding $\prec$. Let $(L_1, \overline{L_1}), \ldots, (L_n, \overline{L_n})$ be the pairs on which A and B conclusion-conclusion attack, where for $i = 1 \ldots n$, L_i and $\overline{L}_i$ are conclusions in A and B respectively. Then A is preferred$_\prec$ to B if for $i = 1 \ldots n$, $R_{L_i}(A + S_i) > R_{\overline{L}_i}(B + S'_i)$

Example 11 In Example 8, $B3$ concludes $r2 \prec r1, r4 \prec r3$, and so $R_a(A3) > R_{\neg a}(A4)$, $R_b(A3) > R_{\neg b}(A4)$, and $A3$ is preferred$_\prec$ to $A4$. In Example 9, $B1$ concludes $r3 \prec r1$, and so $R_a(A1) > R_{\neg a}(A3)$ and $A1$ is preferred$_\prec$ to $A3$.

Observe that:

1. If A and B are strict arguments that conclusion-conclusion attack (recall that they cannot undercut since strict rules contain no weak literals) then the sets of relevant defeasible rules for the pairs of conclusions on which they attack are obviously empty. Hence, it cannot be the case that A is preferred$_\prec$ to B, or B is preferred$_\prec$ to A, for any ordering $\prec$.

2. By Definition 29, $R' > R$ whenever R' is empty and R is non-empty. Hence, if A is strict and B defeasible, and these arguments conclusion-conclusion attack, then for any ordering $\prec$, A is preferred$_\prec$ to B.

7.3 Evaluating the Status of ALP-DP Arguments in Extended Argumentation Frameworks

Prior to instantiating an *EAF* with ALP-DP arguments, we review [34]'s definition of the acceptability of arguments and the grounded extension of an ALP-DP theory:

Definition 31 Let (S, D) be an *ALP-DP* theory. [34] defines:

1. $Args$ to be the set of arguments as defined in Definition 25-1

2. $A1 \in Args$ conclusion-conclusion or undercuts attacks $A2 \in Args$ as defined in Definition 27.

3. $\forall T \subseteq Args$, $\prec_T$ = $\{r \prec s | \ r \prec s$ is the conclusion of an argument in T $\}$.
   $A \in Args$ is preferred$_{\prec_T}$ to $B \in Args$ as defined in Definition 30, except that $\prec_T$ is defined as above rather than by a single priority argument.

4. Let $T \subseteq Args$. Then $A \in Args$ T-defeats $B \in Args$ iff A is the empty argument and B is incoherent (attacks itself), or else if
   (a) A and B conclusion-conclusion attack and B does not undercut A and B is not preferred$_{\prec_T}$ to A; or
   (b) A undercuts B

5. A strictly T-defeats B if A T-defeats B and B does not T-defeat A.

6. $A \in Args$ is ALP-DP-acceptable w.r.t $T \subseteq Args$ if all arguments that T-defeat A are strictly T-defeated by an argument in T.

7. Let $T \subseteq Args$ be conflict free if no argument in T attacks an argument in T, and let 2^{ArgsC} be the set of all conflict free subsets of $Args$. Then:

   - $G : 2^{ArgsC} \mapsto 2^{Args}$, $G(T)$ = $\{A | A$ is ALP-DP-acceptable w.r.t $T\}$
   - G^0 = $\emptyset$, G^{i+1} = $G(G^i)$

8. If (S, D) is finitary, i.e., each argument in $Args$ is attacked by at most a finite number of arguments, then the least fixed point of G (the grounded extension of (S, D)) is obtained by $\bigcup_{i=0}^{\infty}(G^i)$.

Observe that:

1. [34] obviate against the malign impact of incoherent arguments[5] by having them defeated by the empty argument [ ]. We will adopt the same approach when instantiating an *EAF*. Thus, $\mathcal{R}$ will be defined by the set of pairs (A, B) where A undercuts or conclusion-conclusion attacks B, or where B is incoherent, in which case A = [ ].

2. If A undercut attacks B then A defeats B irrespective of their relative preference[6]. Hence, only conclusion-conclusion attacks will be subject to *d*-attack by preference arguments.

3. If A does not undercut B, A conclusion-conclusion attacks B, and in addition B undercuts A, then A cannot defeat B (irrespective of the relative preference of A and B). The rationale for undercuts overriding conclusion-conclusion attacks appeals to intuitions arising from the modelling of legal examples. The fact that B undercuts A effectively means that the conclusion-conclusion attack from A to B cannot succeed as a defeat. This is modelled by the *d*-attack [ ] $\twoheadrightarrow (A \rightharpoonup B)$, where, as we will show, the empty argument cannot be attacked and therefore defeated by any argument.

Definition 32 The *EAF* ($Args$, $\mathcal{R}$, $\mathcal{D}$) for a theory (S, D) is defined as follows.

1. $Args$ is the set of arguments given by Definition 25.
2. $\mathcal{R}$ = $\mathcal{R}1 \cup \mathcal{R}2 \cup \mathcal{R}3$, where:
   (a) $\mathcal{R}1$ = $\{(A, B)|A$ undercuts $B\}$.
   (b) $\mathcal{R}2$ = $\{(A, B)|A$ conclusion-conclusion attacks $B\}$.
   (c) $\mathcal{R}3$ = $\{$([ ],B)$|(B, B) \in (\mathcal{R}1 \cup \mathcal{R}2)\}$.
3. Let $\mathcal{R}'$ = $\mathcal{R} - (\mathcal{R}1 \cup \mathcal{R}3)$. Then $\forall (A, B) \in \mathcal{R}'$:
   (a) $\forall C \in Args$, if C concludes $\prec$ and B is preferred$_{\prec}$ to A then $(C, (A, B)) \in \mathcal{D}$
   (b) if $(B, A) \in \mathcal{R}1$ then ([ ], $(A, B)) \in \mathcal{D}$

Notice that $\mathcal{R}1$ and $\mathcal{R}2$ may not be disjoint. Hence, $\mathcal{R}'$ excludes (from being subject to *d*-attack by preference arguments) those conclusion-conclusion attacks (A, B) in $\mathcal{R}2$ that are also undercut attacks (A, B) in $\mathcal{R}1$. Also, by Definition 27, the empty argument cannot participate in a conclusion-conclusion or undercut with any argument. This means that $\mathcal{R}3$ is necessarily disjoint from $\mathcal{R}1 \cup \mathcal{R}2$. The following observations are shown to hold in Lemma 11 in the Appendix:

[5] Suppose $\{A1, A2, \ldots, An\}$, where $A1$ alone is incoherent, and no two distinct arguments in the set attack each other. $A1$ cannot be in a conflict free set and so a stable extension. Hence, no stable extension exists, since no argument in $\{A2, \ldots, An\}$ attacks and so defeats $A1$. This is counter-intuitive. It's as if the incoherence of $A1$ has 'infected' the other arguments. However, by letting $A1$ be defeated by the empty argument, then $\{$[ ],$A2, \ldots, An\}$ is now a stable extension

[6] This is justified by intuitions arising in the legal domain. However, in [34], the authors acknowledge that this may not be warranted in other domains

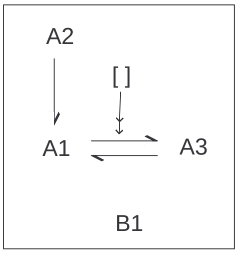
Remark 1 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the *EAF* for (S, D). Let A, $B \in Args$. Then:

1. Δ is preference symmetric.
2. If A undercuts B, then A defeats$_S$ B for any $S \subseteq Args$.
3. If ([ ],(X,Y)) $\in \mathcal{D}$ then there does not exist a (Z,(Y,X)) $\in \mathcal{D}$.
4. ([ ],B) $\in \mathcal{R}$ iff B is incoherent.
5. There is no B such that (B,[ ]) $\in \mathcal{R}$
6. If A does not undercut B, A conclusion-conclusion attacks B and B undercuts A, then A does not defeat$_S$ B for any $S \subseteq Args$ such that S contains [ ].
7. If $A \in Args$ is incoherent, then A is not acceptable w.r.t. any $S \subseteq Args$.

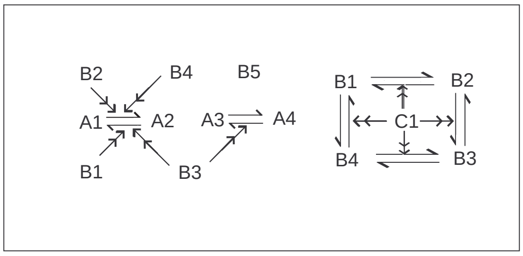
Since the instantiated *EAF* is preference symmetric, its extensions can be obtained on the basis of *ps* acceptability. Note that the requirement on *EAF*s that if (C,(A, B)), (C',(B, A)) $\in \mathcal{D}$ then (C, C'), $(C', C) \in \mathcal{R}$, is satisfied by Definition 32's instantiation of an *EAF*, given that:

Proposition 14 Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the *EAF* for a theory (S, D). If (C,(A, B)), (C',(B, A)) $\in \mathcal{D}$ then C and C' conclusion-conclusion attack.

Example 12 For Example 9, we obtain the *ps-EAF* shown in Figure 8a), in which $B1$ cannot *d*-attack the attack from $A3$ to $A1$ since $A3$ undercuts $A1$. The undercut also means that the empty argument *d*-attacks the conclusion-conclusion attack from $A1$ to $A3$. $\{A2, A3, B1, []\}$ is the single preferred/complete/stable/grounded extension.



a)



b)

Figure 8: a) shows the *EAF* for example 9; b) shows the *EAF* for example 8

Example 13 For Example 8 we obtain the *EAF* in Figure 8b). $\{C1, B1, B3, B5, A1, A3\}$ is a subset of the single grounded, preferred, complete and stable extension E.

In general we can show that:

Proposition 15 Let Δ be the *EAF* for the ALP-DP theory (S, D). Then for $s \in$ {grounded, complete, stable, preferred}, if A is sceptically, respectively credulously, justified under the s semantics, then all sub-arguments of A are sceptically, respectively credulously, justified under the s semantics.

The definition of ALP-DP acceptability in [34] (given here in Definition 31-6) requires that a reinstating argument strictly defeat, and thus precludes definition of admissible and preferred extensions. To see why, suppose one defined an admissible extension as a conflict free set T whose members are ALP-DP acceptable w.r.t. T. Consider then the theory (S, D) where $D = \{r1 :\Rightarrow a, r2 :\Rightarrow \neg a\}$, and $A = [r1]$, $B = [r2]$. A is not ALP-DP acceptable with respect to $\{A\}$ since A does not strictly T-defeat B. Also, B is not ALP-DP acceptable with respect to $\{B\}$. Hence neither can be defined to be admissible (or preferred). Both the instantiated Dung framework ($A \rightleftharpoons B$) and the instantiated *EAF* both return $\{A\}$ and $\{B\}$ as admissible and preferred extensions. Suppose ALP-DP-acceptability was weakened to:

> $A \in Args$ is ALP-DP**2** acceptable w.r.t $T \subseteq Args$ if all arguments that T-defeat A are T-defeated by an argument in T.

Then, if in addition to $r1$ and $r2$ the aforementioned theory also contained $r3 :\Rightarrow r1 \prec r2$, A and $C = [r3]$ would be ALP-DP**2** acceptable with respect to $\{A\}$, but A would not be ALP-DP**2** acceptable with respect to $\{A, C\}$, i.e., $\{A, C\}$ would not be admissible. That is, Dung's fundamental lemma would not hold. As already discussed in Section 4.1, one would want that any definition of acceptability ensure satisfaction of the fundamental lemma. Also, the properties of admissible and preferred extensions described in Proposition 4 and all that is implied by this proposition (including Corollary 1) would not hold. By instantiating an *EAF*, our more of acceptability provides for a well defined notion of an ALP-DP theory's admissible extensions and the justified status of ALP-DP arguments under the preferred semantics.

Example 14 Consider a theory's arguments:
$A = [guardian :\Rightarrow sky \prec bbc]$, $B = [sun :\Rightarrow bbc \prec sky]$, $C = [bbc :\Rightarrow sun \prec guardian]$, $D = [sky :\Rightarrow guardian \prec sun]$, where $x :\Rightarrow y \prec z$ is interpreted as x is the source for the claim that y is less trustworthy than z.

We obtain the non hierarchical *EAF* in Example 3 (see Figure 4). Recall the observation that no two arguments defeat$_S$ each other in $S = \{A, C, B, D\}$. Hence, if we had defined a conflict free set as one in which no two arguments defeat each other, then $\{A, C, B, D\}$ would be the single preferred extension. This strikes us as counterintuitive; that one should simultaneously hold as justified, arguments that claim contradictory preferences. As it is, we have that $\emptyset$ is the grounded extension, and $\{A, C\}$, $\{B, D\}$ the preferred extensions, each of which represent the mutually supportive media outlets; we can either adopt a 'sky and sun' perspective on the world, or a 'guardian and bbc' perspective on the world.

We conclude with a result relating the grounded extension of an ALP-DP theory as defined in [34], to the grounded extension of the theory's *EAF*. In [34], the grounded extension of a finitary ALP-DP theory is the least fixed point of the theory's characteristic function (as described in Definition 31-8). However, a finitary ALP-DP theory does not necessarily yield a finitary *EAF* (recall the definition of a finitary *EAF* in Definition 12). It may be that in the theory's *EAF*, an attack between two conclusion-conclusion attacking arguments may be *d*-attacked by an infinite number of arguments.

Example 15 Let D in (S, D) be the infinite set of defeasible rules:
$\{\ m : \Rightarrow a, n : \Rightarrow \neg a, p : \Rightarrow m \prec n\ \} \cup \{r_X : \Rightarrow r_{X+1} \prec r_{X+2}\ |X$ is a positive

integer }
Let A = $[m]$, B = $[n]$, C_0 = $[p]$, and C_1 = C_0 + $[r_1]$, C_2 = C_0 + $[r_2]$,. . ..
The theory's non-finitary *EAF* is $A \rightleftharpoons B$, where for $i = 0, 1, 2, \ldots$, $C_i \twoheadrightarrow (A \rightharpoonup B)$

Hence, we will only consider ALP-DP theories that contain a finite set of rules deriving a priority relation.

Definition 33 (S, D) is priority-finitary iff

- $\{r | r \in D$, head$(r) = r' \prec r''$ $\}$ is finite
- Any argument in $Args$ is attacked by at most a finite number of arguments in $Args$ (where $Args$ and attack are defined in Definitions 25 and 27 respectively).

Proposition 16 If (S, D) is priority-finitary then its *EAF* Δ = $(Args$, $\mathcal{R}$, $\mathcal{D})$ is finitary.

Proposition 17 states that the grounded extension of a priority-finitary ALP-DP theory as defined in [34], is the same, *modulo the composite priority arguments*, as the grounded extension of the theory's *EAF*. That is to say, the grounded extension of the theory's *EAF* differs only in the contained composite priority arguments.

Definition 34 The sets of arguments S and S' are said to be *equivalent modulo composite priority arguments*, iff:

1. $S' \subseteq S$
2. if $A \in S$, $A \notin S'$ then A is a composite priority argument constructed from the singleton priority arguments $A_1, \ldots, A_n$, where $n > 1$, and for $i = 1 \ldots n$, $A_i \in S'$ (and so $A_i \in S$ given 1.).

Proposition 17 Let (S, D) be a priority-finitary ALP-DP theory, and let GE((S, D)) be the grounded extension of (S, D) as defined in Definition 31-8. Let Δ = $(Args$, $\mathcal{R}$, $\mathcal{D})$ be the theory's *EAF* and GE(Δ) the grounded extension of Δ. Then GE(Δ) and GE((S, D)) are equivalent modulo the composite priority arguments.

8 Related Work

The formalisms of Prakken and Sartor [34] and Kakas and Moraitis [24], both construct arguments from logic programming rules, where these rules may express priorities over other rules. Our work can be viewed as a generalisation of these approaches in that we abstract from the underlying logic and the criteria used for determining preferences. Furthermore, these formalisms do not straightforwardly conform to application of Dung's acceptability calculus, whereas our approach allows for such application, through provision of an abstract characterisation of preference application while retaining the basic conceptual machinery of Dung. We have already shown how arguments defined by an ALP-DP theory instantiate a preference symmetric framework, so that unlike [34], one can define extensions under the admissible and preferred semantics.

In [24], arguments are built from a partitioned theory $(\mathcal{T}, \mathcal{P})$ consisting of logic programming rules with explicit negation, but without negation as failure. $\mathcal{P}$ denotes the set of rules with heads that are of the form $hp(r, r')$ expressing that r has higher priority than r', where r and r' name rules in $\mathcal{T} \cup \mathcal{P}$. An argument is a pair (T, P) such that $T \subseteq \mathcal{T}$, $P \subseteq \mathcal{P}$ and (T, P) derives (by application of modus ponens) some literal L ($T \cup P \vdash L$) (notice that (T, P) may derive more than one literal, and so may contain multiple sets of rules chained through modus ponens). Then, (T, P) symmetrically conflicts with (T', P') iff $T \cup P \vdash L$ and $T' \cup P' \vdash \neg L$ (where L and $\neg L$ may be contradictory priorities on rules). It is the attack relation that accounts for the priorities over rules, so that (T', P') attacks (T, P) iff they conflict on some L and $\neg L$, and such that the minimal subset $T_1 \cup P_1$ of $T' \cup P'$ used to derive L is preferred to (in the sense of being 'at least as strong as') the minimal subset $T_2 \cup P_2$ of $T \cup P$ used to derive $\neg L$. We refer the reader to [24] for how such a preference is defined; suffice it to say that the preference for (T_1, P_1) over (T_2, P_2) is based on the (T', P') derived relative priorities of rules in (T_1, P_1) and (T_2, P_2). Let us here isolate the 'preference argument' (T'_p, P'_p) ($T'_p \subseteq T'$, $P'_p \subseteq P'$) that minimally derives these priorities.

[24] defines the admissibility of an *argument* (T, P):

> (T, P) is admissible iff it is conflict free, and for any (T', P') that attacks (T, P), (T, P) attacks (T', P').

This definition equates with our definition of admissibility. Informally, one can see how to instantiate a preference symmetric *EAF*: by enumerating all $(\mathcal{T}, \mathcal{P})$ defined arguments that are only comprised of single sets of chained rules (as in [34]); identifying preference arguments; defining $\mathcal{R}$ in terms of [24]'s symmetric conflict relation, and; defining $\mathcal{D}$ attacks from preference arguments to $\mathcal{R}$ attacks, corresponding to [24]'s definition of attack. Then, one can see that (T, P) is admissible in [24] iff S = $\{(T_1, P_1), \ldots, (T_n, P_n)\}$ is admissible in the instantiated EAF (where $T = \bigcup_{i=1}^{n} T_i$, $P = \bigcup_{i=1}^{n} P_i$). Notice, that acceptability of a single argument with respect to set of arguments (and so articulation of a Dung style acceptability calculus and characteristic function) is not defined in [24]. Acceptability would need to be derived from the definition of admissibility, i.e., (T, P) is acceptable w.r.t. (T', P') if $(T' \cup T, P' \cup P)$ is admissible. This imposes stronger requirements than our definitions of acceptability and *ps* acceptability, neither of which would require showing that *all* arguments in $(T' \cup T, P' \cup P)$ are acceptable w.r.t. $(T' \cup T, P' \cup P)$ (as highlighted at the end of Section 4.1).

Other non-monotonic formalisms that enable reasoning about preferences, but that do not employ concepts from argumentation theory, include [14] and [17]. For example, [17] make use of a preference operator $\otimes$, so that a defeasible rule such as $\neg sun\_shining \Rightarrow raining \otimes snowing$ expresses that if it is believed that the sun is not shining, then it is believed that it is raining; but if it is not raining then it is believed that is snowing. Given that the sun is not shining, one can see how the defeasible rule could be used in the construction of an argument $A1$ claiming that it is raining, and a symmetrically attacking argument $A2$ claiming that it is snowing, and an argument expressing a default preference for $A1$ over $A2$. If it turns out to be the case that it is not raining, this would constitute an argument defeating $A1$, so that $A2$ would now be reinstated.

We conclude by mentioning the work of [10], in which argument frameworks are similarly extended to allow attacks on attacks. In this way, the strength of one argument's attack on another can be modified by another argument. The authors then describe how the strengths of arguments can propagate through the framework. However, this work does not address the issue of how to define the status of arguments under Dung's semantics.

9 Conclusions

This paper has extended Dung's theory of argumentation to integrate metalevel argumentation about preferences. Dung's level of abstraction is preserved, so that arguments expressing preferences are distinguished by being the source of a second attack relation that abstractly characterises application of preferences by attacking attacks between the arguments that are the subject of the preference claims. Dung's acceptability calculus has been extended to additionally account for reinstatement of attacks, and extensional semantics are then (apart from the grounded semantics) defined in exactly the same way as for a Dung framework. The extended theory thus retains the basic conceptual machinery of Dung's theory, and some basic results that hold for a Dung framework are also shown to hold for the extended framework. We therefore propose that applications and developments of Dung's theory can be readily modified to accommodate argumentation about preferences. In particular, proof theories [30] to determine the justified status of an argument A, take the form of games played between the proponent and opponent of A. Each player responds to its counterpart's move (argument) with an argument that attacks its counterpart's argument, and rules on the legality of moves varying according to the semantics under which the justified status of A is to be established. Ongoing work is focussing on adapting these proof theories to *EAF*s, by additionally allowing players to move arguments against the attacks that are implicitly moved by their counterparts.

Future work will also focus on development of dialogue frameworks in which participants can argue about preferences. Dung's inherently dialectical theory, and subsequent development of argument games, has informed development of dialogue frameworks in which participants' locutions implicitly define arguments that can be organised into a Dung framework. For example, Prakken defines a general framework for conflict resolution dialogues [33] that we believe can readily be generalised to accommodate preference arguments that can then be organised into an *EAF*. In a recent general framework for negotiation dialogues [4], an agent can decide to accept or reject offers based on its own argumentation based model of reasoning about offers. The model assumes a *given* preference ordering on arguments. However, a comprehensive account of negotiation will also require that agents can justify and debate their preferences. Work on deliberation dialogues [21], in which agents collaboratively decide on a preferred course of action, have also highlighted the need to debate the preferences that each agent uses when proposing or rejecting a course of action.

Our work provides a formal setting for studying application of preferences, and more generally, removal of attacks. We have shown that a less involved definition of acceptability suffices for preference symmetric *EAF*s, and that while in general an *EAF*'s characteristic function is not monotonic, the characteristic functions of hierarchical and preference symmetric *EAF*s do satisfy monotonicity. Our work on preference symmetric *EAF*s can also inform work on *resolution* based semantics [9, 26], whereby a framework's resolution obtained by substituting asymmetric for symmetric attacks, and indeed the choice of resolution, can now be explicitly modelled through integration of preference arguments.

The extended theory is proposed as a unifying formalism for works that augment Dung's abstract theory with preferences, and as a framework for instantiation by non-monotonic logics whose object level languages allow for expression of priorities over rules. In Section 5 we described how value based argumentation [11], extended to allow integration of meta-level reasoning about values and value orderings, can be formalised in hierarchical *EAF*s. In Section 7, the inferences of an ALP-DP theory

[34] were shown to correspond with the claims of arguments in the grounded extension of an *EAF* instantiated by the theory's arguments and attacks. In contrast with [34], we also provided a well defined notion of the admissible and preferred extensions of an ALP-DP theory. It remains to formally show that extended argumentation semantics can be given for other object level formalisms such as [14, 17, 24].

In Section 5 we described a practical reasoning example that assumed arguments constructed in an underlying BDI logic as described in [8]. This work is one amongst a number of recent works (e.g., [24, 25, 35]) that propose argumentation based approaches to agent reasoning over the whole gamut of mental attitudes, including beliefs, goals, intentions, e.t.c. We suggest that the framework described in this paper will facilitate development of agent reasoning formalisms that provide for defeasible reasoning about preferences, and thus provide for agent flexibility and adaptability (this proposal is explored in more detail in [29, 31] and also by Kakas and Moraitis [23, 24]). For example, the BOID agent architecture characterises generated candidate goal sets as extensions of a prioritised default logic theory in which rules for inferring goals are modelled as defaults, and a prioritisation of these defaults resolves conflicts between goals [22]. Agent behavioural types correspond to different prioritisations. For example, a 'social' agent uniformly prioritises defaults for obligation derived goals above defaults for desire derived goals. However, certain contexts may warrant selfish behavior (corresponding to the reverse prioritisation). Such behavioural heterogeneity requires reasoning as to which prioritisation (agent type) is appropriate in a given context. In [29], we suggest that extended argumentation frameworks will facilitate development of logical formalisms for contextual reasoning of this type.

Acknowledgements This work was funded by the EU FP6-IST-002307 ASPIC project. I would like to thank Martin Caminada for early discussions that inspired development of the ideas in the paper. Many thanks also to Trevor Bench-Capon and P.M. Dung for valuable discussions and comments that have helped improve the paper. Finally, thanks to the anonymous reviewers whose comments helped improve the paper.

10 Appendix

10.1 Proofs of Propositions in Section 3

Proposition 1 If A defeats$_S$ B then $\forall S' \subseteq S$, A defeats$_{S'}$ B.

Proof: Follows straightforwardly from Definition 5.

Proposition 2 Let S be conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then for any $X, Y \in S$, X does not defeat$_S$ Y.

Proof: Suppose $X, Y \in S$, $X \rightarrow^S Y$. In which case $(X, Y) \in \mathcal{R}$ and $\neg\exists Z \in S$ s.t. (Z,(X, Y)) $\in \mathcal{D}$, contradicting S is conflict free.

10.2 Proofs of Propositions in Section 4

Lemma 1 is required for the proof of Proposition 3.

Lemma 1 Let S be a conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$), and let A be an argument that is acceptable w.r.t. S. Then:

1. A does not defeat$_S$ A

2. $\forall C \in S$, C does not defeat$_S$ A

Proof: **(1)** Suppose $A \rightarrow^S A$. By assumption of A acceptable w.r.t. S, $\exists D \in S$ s.t. $D \rightarrow^S A$, $\exists E \in S$ s.t. $E \rightarrow^S D$, contradicting Proposition 2. **(2)** Suppose $\exists C \in S$ s.t. $C \rightarrow^S A$. By assumption of A acceptable w.r.t. S, $\exists D \in S$ s.t. $D \rightarrow^S C$, contradicting Proposition 2.

The following lemma and proof of Proposition 3 refer to an *EAF*s characteristic function defined in Definition 10.

Lemma 2 Let S and S' be subsets of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). If $S \subseteq S'$ and no argument in S defeats$_S$ an argument in $S' - S$. Then $F(S) \subseteq F(S')$.

Proof: Let $A \in F(S)$. Suppose $\exists B$ s.t. $B \rightarrow^{S'} A$. By Proposition 1, $B \rightarrow^S A$, and so $\exists C$, $C \rightarrow^S B$ and there is a reinstatement set R_S for $C \rightarrow^S B$. We show that $C \rightarrow^{S'} B$ and there is a reinstatement set $R_{S'}$ for $C \rightarrow^{S'} B$:
Suppose otherwise. In which case $X \not\rightarrow^{S'} Y$ for some $X \rightarrow^S Y \in R_S$, i.e., $X \rightharpoonup Y$ and $\exists B' \in S' - S$ s.t. $B' \twoheadrightarrow (X \rightharpoonup Y)$. But then by assumption of R_S being a reinstatement set, $\exists C' \in S$ s.t. $C' \rightarrow^S B'$, contradicting the assumption that no argument in S defeats$_S$ an argument in $S' - S$.

Proposition 3 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an Extended Argumentation Framework, S an admissible extension of Δ, and let A, A' be arguments which are acceptable w.r.t S. Then:

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

Proof:
1) By assumption of A acceptable w.r.t S and Lemma 1-2, no argument in S defeats$_S$ an argument in $S' - S$ (= $\{A\}$). Hence, by Lemma 2, $F(S) \subseteq F(S')$. Since S is admissible and A is acceptable w.r.t. S, then $S \cup \{A\} \subseteq F(S)$, and so $S' \subseteq F(S')$, i.e., S' is admissible.
2) Follows from $F(S) \subseteq F(S')$ and assumption of A' acceptable w.r.t. S.

Proposition 4 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an Extended Argumentation Framework.

1. The set of all admissible extensions of Δ form a complete partial order w.r.t. set inclusion
2. For each admissible extension E of Δ there exists a preferred extension E' of Δ such that $E \subseteq E'$

Proof: Immediately from Proposition 3 and definition of preferred extensions (Definition 9).

Corollary 1 Every *EAF* possesses at least one preferred extension.

Proof: From Proposition 4 and the fact that $\emptyset$ is an admissible extension of any *EAF*.

Proposition 5 Every stable extension of an *EAF* is a preferred extension but not vice versa.

Proof: It is clear that each stable extension is a preferred extension. To show the reverse does not hold, one need only consider an *EAF* containing a single argument that attacks itself. $\emptyset$ is the single preferred extension. There is no stable extension.

We establish some lemmas required for the proofs of the propositions that follow.

Lemma 3 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be an *EAF*. If (X,Y),$(Y,X) \in \mathcal{R}$, then for any conflict free $S \subseteq Args$, either $X \rightharpoondown^S Y$ or $Y \rightharpoondown^S X$.

Proof: Suppose otherwise, i.e., $X \not\rightharpoondown^S Y$ and $Y \not\rightharpoondown^S X$. In which case $\exists Z, Z' \in S$ s.t. $Z \twoheadrightarrow (X \rightharpoonup Y)$, $Z' \twoheadrightarrow (Y \rightharpoonup X)$, and so by definition of an *EAF*, $Z \rightleftharpoons Z'$, contradicting S is conflict free.

Lemma 4 Let S be a conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$), and A, B be arguments such that A defeats$_S$ B. Then it cannot be that both A and B are acceptable w.r.t. S.

Proof: Suppose B is acceptable w.r.t. S. Then $\exists C \in S$ s.t. $C \rightharpoondown^S A$. Suppose A is acceptable w.r.t. S. Then $\exists C' \in S$ s.t. $C' \rightharpoondown^S C$, contradicting Proposition 2

Lemma 5 Let S be a conflict free subset of $Args$ in ($Args$, $\mathcal{R}$, $\mathcal{D}$). Suppose $S \subseteq F(S)$. Then $F(S)$ is conflict free.

Proof: Assume $F(S)$ is not conflict free. Then $\exists A, B \in F(S)$ such that either

1. $(A,B) \in \mathcal{R}$, $(B,A) \notin \mathcal{R}$, $\neg\exists C \in F(S)$ s.t. $(C,(A,B)) \in \mathcal{D}$. Since $S \subseteq F(S)$, then $A \rightharpoondown^S B$; or
2. $(A,B) \in \mathcal{R}$, $(B,A) \in \mathcal{R}$. By Lemma 3, either $A \rightharpoondown^S B$ or $B \rightharpoondown^S A$.

In both cases $\exists X, Y \in F(S)$ s.t. X defeats$_S$ Y, contradicting Lemma 4.

Proposition 6 is a special case of Lemma 6:

Lemma 6 Let F be the characteristic function of ($Args$, $\mathcal{R}$, $\mathcal{D}$), F^0_+ any subset of the set $\{A|A$ is acceptable w.r.t. $\emptyset\}$, and $F^{i+1}_+ = F(F^i_+)$. Then $\forall i$, $F^i_+ \subseteq F^{i+1}_+$ and F^i_+ is conflict free.

Proof:
Base case: i = *0,1*. If $(B,A) \in \mathcal{R}$, then by Definition 5, it must be that $B \rightharpoondown^{\emptyset} A$ and so A cannot be acceptable w.r.t. $\emptyset$. Hence, F^0_+ is some subset of $\{A|\forall B \in Args, (B,A) \notin \mathcal{R}\}$, i.e., some subset of the set of arguments that are not defeated$_S$ for any $S \subseteq Args$. Hence, $F^0_+ \subseteq F^1_+$ and F^0_+ is conflict free.
By Lemma 5, $F^1_+ = F(F^0_+)$ is conflict free. By Lemma 1-2 no argument in F^0_+ defeats$_{F^0_+}$ an argument in F^1_+. Hence, by Lemma 2, $F(F^0_+) \subseteq F(F^1_+)$, i.e., $F^1_+ \subseteq F^2_+$.

Inductive hypothesis: For $j < i$, $F^j_+ \subseteq F^{j+1}_+$ and F^j_+ is conflict free.

General case: i. By inductive hypothesis, F^{i-1}_+ is conflict free and $F^{i-1}_+ \subseteq F^i_+$, and so since $F^i_+ = F(F^{i-1}_+)$, then by Lemma 5 F^i_+ is conflict free. Since F^{i-1}_+ is conflict free and $F^i_+ = F(F^{i-1}_+)$, then by Lemma 1-2 no argument in F^{i-1}_+ defeats$_{F^{i-1}_+}$ an argument in F^i_+. Given $F^{i-1}_+ \subseteq F^i_+$, then by Lemma 2, $F(F^{i-1}_+) \subseteq F(F^i_+)$. That is, $F^i_+ \subseteq F^{i+1}_+$.

Proposition 6 Let F be the characteristic function of an *EAF*, and $F^0 = \emptyset$, $F^{i+1} = F(F^i)$. Then $\forall i$, $F^i \subseteq F^{i+1}$ and F^i is conflict free.

Proof: Follows as a special case of Lemma 6, where $F^0_+ = F^0 = \emptyset$.

10.3 Proofs of Propositions in Section 5

Proposition 7 Let F be the characteristic function of a bounded hierarchical *EAF* Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$). Let S and S' be conflict free subsets of $Args$ such that $S \subseteq S'$. Then $F(S) \subseteq F(S')$.

Proof: Assume A is acceptable w.r.t. S. We show that A is acceptable w.r.t. S'

Suppose $\exists B1$ s.t. $B1 \rightarrow^{S'} A$. By Proposition 1, $B1 \rightarrow^{S} A$, and by assumption of A acceptable w.r.t. S, $\exists C1 \in S$ s.t. $C1 \rightarrow^{S} B1$ and there is a reinstatement set R_S for $C1 \rightarrow^{S} B1$.

Since $S \subseteq S'$, $C1 \in S'$, and so to show A is acceptable w.r.t. S', it is sufficient to show there is a reinstatement set $R_{S'}$ for $C1 \rightarrow^{S'} B1$.

Since Δ is hierarchical, we can identify a sequenced partition of the reinstatement set R_S for $C1 \rightarrow^{S} B1$. Let Δ_H = ($((Args_1, \mathcal{R}_1), \mathcal{D}_1)$, ..., $((Args_n, \mathcal{R}_n), \mathcal{D}_n)$) be the partition of Δ. Since Δ is bounded ($\mathcal{D}_n$ = $\emptyset$) one can represent R_S by the finite:

$$R_{S_i} = \{C1 \rightarrow^{S} B1\} \cup R_{S_{i+1}} \cup \ldots \cup R_{S_k}, \text{ where:}$$

1) for each defeat in R_{S_j}, $j < k$, any defence attack on the defeat originates from an argument that is itself defeated in $R_{S_{j+1}}$. That is, for $j = i \ldots k-1$:
R_{S_j} = $\{C_1^j \rightarrow^S B_1^j, C_2^j \rightarrow^S B_2^j \ldots\}$ such that for $m = 1, 2, \ldots$, $(C_m^j, B_m^j) \in \mathcal{R}_j$, and if $B \twoheadrightarrow (C_m^j \rightharpoonup B_m^j)$ then B is some $B^{j+1} \in Args_{j+1}$, and $C^{j+1} \rightarrow^S B^{j+1} \in R_{S_{j+1}}$

2) since Δ is bounded, R_{S_k} is a set of defeats $\{C_1^k \rightarrow^S B_1^k, C_2^k \rightarrow^S B_2^k, \ldots\}$ such that for $m = 1, 2, \ldots$, $\neg\exists\ B \in Args$ s.t. $B \twoheadrightarrow (C_m^k \rightharpoonup B_m^k)$.

We show the existence of $R_{S'}$ by showing that if $C \rightarrow^S B \in R_S$ then $C \rightarrow^{S'} B$. Proof is by induction on the above sequenced partition:

Base case: $C \rightarrow^S B \in R_{S_k}$. Suppose $C \not\rightarrow^{S'} B$. Then $\exists B' \in S'$, $B' \twoheadrightarrow (C \rightharpoonup B)$, contradicting **2)**.

Inductive hypothesis: For $l > j$, $C \rightarrow^S B \in R_{S_l}$ implies $C \rightarrow^{S'} B$.

General case: $C \rightarrow^S B \in R_{S_j}$. Suppose $C \not\rightarrow^{S'} B$. Then $\exists B' \in S'$, $B' \twoheadrightarrow (C \rightharpoonup B)$. Since R_S is a reinstatement set, $\exists C' \in S$, $C' \in Args_{j+1}$, $C' \rightarrow^S B' \in R_{S_{j+1}}$. By inductive hypothesis, $C' \rightarrow^{S'} B'$. But then since $C', B' \in S'$ and S' is conflict free by assumption, this contradicts Proposition 2.

Lemma 7 states that the characteristic function of a finitary *EAF* is ω-continuous. This is required for the proof of Proposition 8.

Lemma 7 Let F be the characteristic function of a finitary *EAF* Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$). Then F is ω-continuous.

Proof: Let $S_0 \subseteq \ldots \subseteq S_n \subseteq \ldots$ be an increasing sequence of sets of arguments, and let $S = S_0 \cup \ldots \cup S_n \cup \ldots$. Let $A \in F(S)$. Since there are finitely many arguments B that attack and so defeat$_S$ A, and for any such $B \rightarrow^S A$, there are finitely many arguments C s.t. $(C,(B,A)) \in \mathcal{D}$, then there exists a number m s.t. $A \in F(S_m)$. Therefore, $F(S) = F(S_0) \cup \ldots \cup F(S_n) \cup \ldots$

Proposition 8 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be a bounded hierarchical *EAF*. Then:

1. $\bigcup_{i=0}^{\infty}(F^i) \subseteq GE(\Delta)$
2. If Δ is finitary, then $\bigcup_{i=0}^{\infty}(F^i) = GE(\Delta)$

Proof:

1. Immediately from the monotonicity of F for bounded hierarchical *EAF*s (Proposition 7)
2. From Lemma 7

Proposition 9 Let $\Gamma = \langle Args,\mathcal{R},V,val,a\rangle$ be an $aVAF$ and Δ its equivalent EAF. Let $s \in \{$preferred, grounded, complete$\}$. Then $\forall A \in Args$, A is a sceptically, respectively credulously, justified argument of Γ under the s semantics, iff A is a sceptically, respectively credulously, justified argument of Δ under the s semantics.

Proof: Let $(Args, \mathcal{R}_a)$ be obtained on the basis of Γ as defined in Definition 17. Let $\Delta = (Args_\Delta, \mathcal{R}_\Delta, \mathcal{D})$. Then $\forall Z, Y \in Args$ s.t. $val(Z) = v$, $val(Y) = v'$:
$(Z, Y) \in \mathcal{R}_a$ iff

- $Args_\Delta$ contains the arguments $Z, Y, val(Z) > val(Y), val(Y) > val(Z)$ and a
- $\mathcal{R}_\Delta$ contains the attacks (Z, Y), $(val(Z) > val(Y), val(Y) > val(Z))$, and $(val(Y) > val(Z), val(Z) > val(Y))$
- $\mathcal{D}$ contains the defence attacks $(val(Y) > val(Z),(Z, Y))$ and $(a, (val(Y) > val(Z), val(Z) > val(Y)))$

By definition (for both Dung and extended frameworks), a complete extension is a fixed point of a framework's characteristic function, and the grounded and preferred extensions are, respectively, the least and maximal (under set inclusion) complete extensions. Hence, let F_Γ and F_Δ be the characteristic functions of $(Args, \mathcal{R}_a)$ and Δ respectively. The proposition is proved by showing:
S is a fixed point of F_Γ iff S' is a fixed point of F_Δ, where $S \subseteq S'$ and $(S' - S) \cap Args = \emptyset$, which holds iff:

$\forall Z \in Args$, Z is acceptable w.r.t. S iff Z is acceptable w.r.t. S' **(i)**

Prior to proving **(i)**, note that since $a \in Args_\Delta$ is not attacked by any argument, then:

every fixed point of F_Δ contains a, and $\{v > v' |\ (a, (v' > v, v > v')) \in \mathcal{D}\}$
$(= \{v > v' | v >_a v'\})$. **(ii)**

(i) is shown by showing 1) and 2) below:

1. $\forall Y, X \in Args$, $(Y, X) \in \mathcal{R}_a$ iff $Y \rightarrow^{S'} X$.
   This follows from $(Y, X) \in \mathcal{R}_a$ iff $val(X) \not>_a val(Y)$ iff $val(Y) >_a val(X)$ iff (by **(ii)**) $val(Y) > val(X) \in S'$ and so $val(X) > val(Y) \notin S'$ (since S' is conflict free) and so $Y \rightarrow^{S'} X$.
2. $\forall Z \in S'$, $\forall Y \notin S'$, if $Z \rightarrow^{S'} Y$ then there is a reinstatement set $R_{S'}$ for $Z \rightarrow^{S'} Y$.
   There are two cases to consider:
   a) Suppose $Z, Y \in Args$. Given 1), $val(Z) > val(Y) \in S'$. We have $(val(Y) > val(Z)) \twoheadrightarrow (Z \rightharpoonup Y)$. But then $val(Z) > val(Y) \rightarrow^{S'} val(Y) > val(Z)$, and $\neg\exists a' \in Args_\Delta$ s.t. $a' \twoheadrightarrow (val(Z) > val(Y) \rightharpoonup val(Y) > val(Z))$. Hence $R_{S'} = \{Z \rightarrow^{S'} Y, val(Z) > val(Y) \rightarrow^{S'} val(Y) > val(Z)\}$.
   b) Suppose either $Z \notin Args$ or $Y \notin Args$. Then Z is of the form $v > v'$ and Y is of the form $v' > v$, and by **(ii)**, $v >_a v'$ and since $\neg\exists a' \neq a$, $\{v > v' \rightarrow^{S'} v' > v\}$ is a reinstatement set for $v > v' \rightarrow^{S'} v' > v$.

10.4 Proofs of Propositions in Section 6

Lemma 8 is used in the proof of Proposition 10.

Lemma 8 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be a preference symmetric *EAF*, and $X, Y \in Args$ such that $(X, Y) \in \mathcal{R}$. Then $\forall S \subseteq Args$:

1. If S is conflict free, either $X \rightarrow^S Y$ or $Y \rightarrow^S X$
2. If $X \rightarrow^S Y$, and for some conflict free superset S' of S, $X \not\rightarrow^{S'} Y$, then $Y \rightarrow^S X$
3. If X strictly defeats$_S$ Y, then X strictly defeats$'_S$ Y for all conflict free supersets S' of S

Proof:

1. Assume that for some S, $X \not\rightarrow^S Y$. Hence there is a $Z \in S$ s.t. $Z \twoheadrightarrow (X \rightharpoonup Y)$. Since Δ is preference symmetric, $(Y, X) \in \mathcal{R}$. By Lemma 3, $Y \rightarrow^S X$.
2. Since S' is conflict free and $X \not\rightarrow^{S'} Y$, then given (1), $Y \rightarrow^{S'} X$, and so by Proposition 1, $Y \rightarrow^S X$.
3. Assume that for some S', $X \not\rightarrow^{S'} Y$. Then given (2), $Y \rightarrow^S X$, contradicting X strictly defeats$_S$ Y. Assume $Y \rightarrow^{S'} X$. By Proposition 1, $Y \rightarrow^S X$, contradicting X strictly defeats$_S$ Y.

Proposition 10 Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be a preference symmetric *EAF*. Let S be a conflict free subset of $Args$. Then, every argument in S is *ps* acceptable w.r.t. S iff every argument in S is acceptable w.r.t. S.

Proof:
Right to left half: Let $A \in S$. If $\exists B$ s.t. $B \rightarrow^S A$, then $\exists C \in S$ s.t. $C \rightarrow^S B$ and there is a reinstatement set R_S for $C \rightarrow^S B$, i.e., $\forall Y$ such that $Y \twoheadrightarrow (C \rightharpoonup B)$, there exists a $X \rightarrow^S Y \in R_S$ such that $X \in S$. Hence, by Definition 20, A is *ps* acceptable w.r.t. S.

Left to right half: Let $A \in S$. If $\exists B1$ s.t. $B1 \rightarrow^S A$, then $\exists C1 \in S$ s.t. $C1 \rightarrow^S B1$, and $C1 \rightarrow^S B1$ is reinstated. We show there must be a reinstatement set R_S for $C1 \rightarrow^S B1$ (and so A is acceptable w.r.t. S). Suppose otherwise. Then, there exists at least one set of sequences of the form:

$$\begin{array}{l}\{\\ (B_{n+1_1} \twoheadrightarrow (C_{n_1} \rightharpoonup B_n), C_{n_1} \rightarrow^S B_n \ldots B_2 \twoheadrightarrow (C1 \rightharpoonup B1), C1 \rightarrow^S B1)\\ \vdots \\ (B_{n+1_m} \twoheadrightarrow (C_{n_m} \rightharpoonup B_n), C_{n_m} \rightarrow^S B_n \ldots B_2 \twoheadrightarrow (C1 \rightharpoonup B1), C1 \rightarrow^S B1)\\ \}\end{array} \quad \textbf{(1)}$$

such that:
- for some $n \geqq 2$, $\{C_{n_1} \ldots C_{n_m}\}$ is the set of all arguments in S that defeat$_S$ B_n **(2)**
- for $j = 1 \ldots m$, $\neg\exists C_{n+1_j} \in S$ s.t. $C_{n+1_j} \rightarrow^S B_{n+1_j}$. **(3)**

Intuitively, given a preference argument B_n attacking an attack that originates from an argument in S, then for the set $\{C_{n_1} \rightarrow^S B_n \ldots C_{n_m} \rightarrow^S B_n\}$ of all potential reinstating defeats, each such defeat must itself be challenged and not defeat reinstated (this is visualised in Figure 9). Note that it must be that $n \geqq 2$ since $C1 \rightarrow^S B1$ is reinstated, and so there must at least be a set $\{C_{2_1} \ldots C_{2_m}\}$. Note also that since for $j = 1 \ldots m$, $C_{n_j} \rightarrow^S B_n$ is challenged by a preference argument, then since Δ is

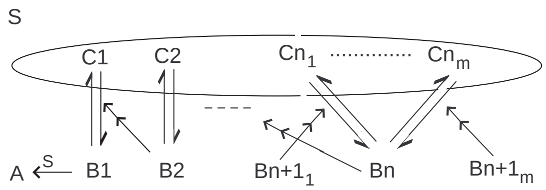


Figure 9: Conditions under which there is no reinstatement set for $C1 \rightarrow_S B1$

preference symmetric, we have that for $j = 1 \ldots m$, $(B_n, C_{n_j}) \in \mathcal{R}$.

Proof is by contradiction. Suppose no reinstatement set R_S for $C1 \rightarrow^S B1$, and so there exists a set of sequences as described in **(1-3)**. There are two cases to consider:

i) Suppose that for some $j = 1 \ldots m$, $B_n \not\rightarrow^S C_{n_j}$. Hence, $\exists B' \in S$ s.t. $B' \twoheadrightarrow (B_n \rightharpoonup C_{n_j})$. By **(1)** and **(2)**, $\exists B_{n+1_j}$, $B_{n+1_j} \twoheadrightarrow (C_{n_j} \rightharpoonup B_n)$ and $C_{n_j} \rightarrow^S B_n$, and so it must be the case that $B_{n+1_j} \notin S$.

Hence, $\exists B' \in S$, $B_{n+1_j} \notin S$, and since they express contradictory preferences, $B' \rightleftharpoons B_{n+1_j}$. Since S is conflict free, then by Lemma 8-1, either $B_{n+1_j} \rightarrow^S B'$ or $B' \rightarrow^S B_{n+1_j}$. If $B' \rightarrow^S B_{n+1_j}$, then this contradicts **(3)**. If $B_{n+1_j} \rightarrow^S B'$, then since $B' \in S$ and by assumption every argument in S is *ps* acceptable w.r.t. S, then there must exist a $C' \in S$ s.t. $C' \rightarrow^S B_{n+1_j}$, contradicting **(3)**.

ii) Suppose that for $j = 1 \ldots m$, $B_n \rightarrow^S C_{n_j}$. Since by **(2)**, $\{C_{n_1} \ldots C_{n_m}\}$ is the set of *all* arguments in S that defeat$_S$ B_n, and by assumption each C_{n_j} is *ps* acceptable w.r.t S, then it must be the case that at least one $C_{n_j} \rightarrow^S B_n$ is reinstated (otherwise no C_{n_j} would be *ps* acceptable). That is, for some B_{n+1_j} there exists a $C' \in S$ s.t. $C' \rightarrow^S B_{n+1_j}$, contradicting **(3)**.

Corollary 2 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be a *ps-EAF*. Let S be an admissible extension of Δ. If $A \in Args$ is *ps* acceptable w.r.t. S then A is acceptable w.r.t. S.

Proof: The proof proceeds in the same way as for the left to right half of Proposition 10, except that $A \in Args$ (rather than restricting to $A \in S$) and in i) and ii) the assumption of arguments in S being *ps* acceptable w.r.t. S is replaced by the assumption of arguments in S being acceptable w.r.t. S.

Proposition 11 Let F be the characteristic function of a preference symmetric *EAF* ($Args$, $\mathcal{R}$, $\mathcal{D}$). Let S and S' be conflict free subsets of $Args$ such that $S \subseteq S'$. Then $F(S) \subseteq F(S')$.

Proof: Given Lemma 2 it suffices to show that no $A \in S$ defeats$_S$ some $B \in S' - S$. Suppose otherwise. Hence $(A, B) \in \mathcal{R}$. Since S' is conflict free, (C,(A, B)) $\in \mathcal{D}$ and $(B, A) \notin \mathcal{R}$, contradicting Δ being a *ps-EAF*.

Proposition 12 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be a preference symmetric *EAF*, F^0 = $\emptyset$, F^{i+1} = $F(F^i)$. Then:

1. $\bigcup_{i=0}^{\infty}(F^i) \subseteq GE(\Delta)$
2. If Δ is finitary, then $\bigcup_{i=0}^{\infty}(F^i)$ = $GE(\Delta)$

Proof:

1. Immediately from the monotonicity of F for preference symmetric *EAF*s (Proposition 11)
2. From Lemma 7

Lemmas 9 and 10 are used in the proof of Proposition 13.

Lemma 9 Let S be an admissible extension of a *ps-EAF* Δ. If A is strict-acceptable w.r.t. S then A is acceptable w.r.t. S.

Proof: Given Corollary 2, it suffices to show that if A is strict-acceptable w.r.t. S then A is *ps* acceptable w.r.t. S. Suppose $\exists B$ s.t. B defeats$_S$ A, and so $\exists C \in S$ s.t. C strictly defeats$_S$ B. We need to show that $C \rightarrow^S B$ is reinstated:

Suppose $\exists B1$ s.t. ($B1$,(C, B)) $\in \mathcal{D}$. Since Δ is a *ps-EAF*, then $B \rightharpoonup C$. By assumption of C strictly defeating$_S$ B, $B \not\rightarrow^S C$, and so $\exists D_1 \in S$ s.t. $(D1, (B, C)) \in \mathcal{D}$. Hence, $D1 \rightleftharpoons B1$. S is conflict free and so (by Lemma 3) $B1 \rightarrow^S D1$ or $D1 \rightarrow^S B1$. If $D1 \rightarrow^S B1$ then $C \rightarrow^S B$ is reinstated. If $D1 \not\rightarrow^S B1$, $B1 \rightarrow^S D1$, then since by assumption of S being admissible, $D_1 \in S$ is acceptable w.r.t. S, and so $\exists C_1 \in S$ s.t. $C_1 \rightarrow^S B1$. Hence $C \rightarrow^S B$ is reinstated.

Lemma 10 Let Δ be a *ps-EAF* ($Args$, $\mathcal{R}$, $\mathcal{D}$) and F^i be defined as in Definition 11. Then $\forall Y \in Args$, if Y is acceptable w.r.t. F^i then Y is strict-acceptable w.r.t. F^i.

Proof:
Base Case By Proposition 6, $\forall i, F^i \subseteq F^{i+1}$. Hence:

BC1 $F^1 = \{A_1, \ldots, A_n\}$ is the set of arguments acceptable w.r.t. $\emptyset$. Hence:

for $i = 1 \ldots n$: $\forall X \in Args$, $(X, A_i) \notin \mathcal{R}$, and so A_i is strict acceptable w.r.t. any $S \subseteq Args$ **(i)**

and since Δ is a *ps-EAF*, $\neg\exists Y$ s.t. $(Y, (A_i, X)) \in \mathcal{D}$, and so:

for $i = 1 \ldots n$: $\forall X \in Args$ s.t. $(A_i, X) \in \mathcal{R}$, $\forall S \subseteq Args$, A_i strictly defeats$_S$ X **(ii)**

BC2 $F^2 = \{B_1, \ldots, B_m, A_1, \ldots, A_n\}$ is the set of arguments acceptable w.r.t. F^1

- **BC2.1** By **(i)**, each A_i is strict acceptable w.r.t. F^1.
- **BC2.2** Suppose for some $i = 1 \ldots m$, $\exists X$ s.t. $X \rightarrow^{F^1} B_i$. By assumption of B_i acceptable w.r.t. F^1, $\exists A_j \in F^1$ s.t. $A_j \rightarrow^{F^1} X$. By **(ii)**, A_j strictly $\rightarrow^{F^1}$ X.

BC3 $F^3 = \{C_1, \ldots, C_k, B_1, \ldots, B_m, A_1, \ldots, A_n\}$ is the set of arguments acceptable w.r.t. F^2

- **BC3.1** By **(i)**, each A_i is strict acceptable w.r.t. F^2.
- **BC3.2** Suppose for some $i = 1 \ldots m$, $\exists X$ s.t. $X \rightarrow^{F^2} B_i$. By Proposition 1, $X \rightarrow^{F^1} B_i$. By assumption of B_i acceptable w.r.t. F^1, $\exists A_j \in F^1$ s.t. $A_j \rightarrow^{F^1} X$. By **(ii)**, A_j strictly $\rightarrow^{F^2}$ X
- **BC3.3** Suppose for some $i = 1 \ldots k$, $\exists X$ s.t. $X \rightarrow^{F^2} C_k$. By assumption of C_k acceptable w.r.t. F^2, $\exists Z \in F^2$ s.t. $Z \rightarrow^{F^2} X$. Suppose $Z \in \{A_1, \ldots, A_n\}$. Then by **(ii)**, Z strictly $\rightarrow^{F^2}$ X. Suppose $Z \in \{B_1, \ldots, B_m\}$, and Z does not strictly $\rightarrow^{F^2}$ X. Then $X \rightarrow^{F^2} Z$. But then **BC3.2** shows that for some $j = 1 \ldots n$, A_j strictly $\rightarrow^{F^2}$ X.

Inductive Hypothesis: For $j < i$, if Y is acceptable w.r.t. F^j then Y is strict-acceptable w.r.t. F^j.

General case: Let Y be any argument acceptable w.r.t. F^i. Suppose $X \rightarrow^{F^i} Y$. Hence, $\exists Z \in F^i$ s.t. $Z \rightarrow^{F^i} X$. Suppose Z does not strictly $\rightarrow^{F^i} X$. Then $X \rightarrow^{F^i} Z$. By Proposition 6, $F^{i-1} \subseteq F^i$, and so by Proposition 1, $X \rightarrow^{F^{i-1}} Z$. Since $Z \in F^i$, then by inductive hypothesis Z is strict-acceptable w.r.t. F^{i-1}; that is, there exists a $Z' \in F^{i-1}$ s.t. Z' strictly $\rightarrow^{F^{i-1}} X$. By Proposition 6, $F^{i-1} \subseteq F^i$, and so $Z' \in F^i$, and by Lemma 8-3, Z' strictly $\rightarrow^{F^i} X$.

Proposition 13 Let Δ be a preference symmetric *EAF*, and let F^i be defined as in Definition 11, $F*^i$ defined as in Definition 23. Then $F*^i = F^i$.

Proof:

Base Case: $F*^0 = F^0 = \emptyset$. $F*^1 = F^1 = \{A | \forall B \in Args, (B, A) \notin \mathcal{R}\}$.
Suppose A is acceptable w.r.t. F^1. By Lemma 10, A is strict-acceptable w.r.t.F^1, and so $F*^1$. Suppose A is strict-acceptable w.r.t. $F*^1$. Since $F*^1 = F^1$,and F^1 is admissible, then by Lemma 9, A is acceptable w.r.t. F^1.

Inductive Hypothesis: For $j < i$, $F*^j = F^j$, and F^j is admissible.

General case: Suppose A is acceptable w.r.t. F^{i-1}. By Lemma 10, A is strict-acceptable w.r.t. F^{i-1} (= $F*^{i-1}$ by inductive hypothesis). Suppose A is strict-acceptable w.r.t. $F*^{i-1}$. By inductive hypothesis, $F*^{i-1} = F^{i-1}$, and F^{i-1} is admissible. By Lemma 9, A is acceptable w.r.t. F^{i-1}.

Corollaries 3 and 4 are used in proofs in Section 7. Corollary 3 restates Proposition 13 showing equivalence of the iterations obtained by F and F_{st} respectively starting with F^0_+ (containing some subset of arguments acceptable w.r.t. $\emptyset$) and $F*^0_+$ (containing some subset of arguments strict acceptable w.r.t. $\emptyset$).

Corollary 3 Let Δ be a preference symmetric *EAF*, and let F^i_+ be defined as in Lemma 6. Let $F*^0_+$ be some subset of $\{A | A$ is strict-acceptable w.r.t. $\emptyset\}$, $F*^{i+1}_+ = F_{st}(F*^i_+)$. Then $F*^i_+ = F^i_+$.

Proof: Proof follows from Lemma 9 and Lemma 10, where:

- 'F^i_+' substitutes for 'F^i' in the statement of Lemma 10
- In the proof of Lemma 10: i) Lemma 6 substitutes for Proposition 6; ii) in proof of the base case, F^0_+, F^1_+ and F^2_+ respectively substitute for F^1, F^2 and F^3; iii) in **BC1**, "$\{A_1, \ldots, A_n\}$ **is some subset** of the arguments $\{A_1, \ldots, A_n\}$ acceptable w.r.t. $\emptyset$" replaces "$\{A_1, \ldots, A_n\}$ is the set of arguments acceptable w.r.t. $\emptyset$"
- F^i_+ substitutes for F^i in the inductive hypothesis and general case.

Corollary 4 Let Δ be a preference symmetric *EAF*, $F*^0_+$ be some subset of $\{A | A$ is strict-acceptable w.r.t. $\emptyset\}$, $F*^{i+1}_+ = F_{st}(F*^i_+)$. Then $\forall i$, $F*^i_+ \subseteq F*^{i+1}_+$ and $F*^i_+$ is conflict free.

Proof: Proof follows immediately from Lemma 6 and Corollary 3.

10.5 Proofs of Propositions in Section 7

Lemma 11 proves results stated in Remark 1.

Lemma 11 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the *EAF* for (S, D), and let A, $B \in Args$. Then:

1. Δ is preference symmetric.
2. If A undercuts B, then A defeats$_S$ B for any $S \subseteq Args$.
3. If ([ ],(X,Y)) $\in \mathcal{D}$ then there does not exist a (Z,(Y,X)) $\in \mathcal{D}$.
4. ([ ],B) $\in \mathcal{R}$ iff B is incoherent.
5. There is no B such that (B,[ ]) $\in \mathcal{R}$
6. If A does not undercut B, A conclusion-conclusion attacks B and B undercuts A, then A does not defeat$_S$ B for any $S \subseteq Args$ such that S contains [ ].
7. If $A \in Args$ is incoherent, then A is not acceptable w.r.t. any $S \subseteq Args$.

Proof:

1. If (C,(A, B)) $\in \mathcal{D}$ as defined in Definition 32-3a then $(A, B) \in \mathcal{R}2$, i.e., A and B conclusion-conclusion attack, and so $(B, A) \in \mathcal{R}2$. If (C,(A, B)) $\in \mathcal{D}$ as defined in Definition 32-3b then $(A, B) \in \mathcal{R}2$, $(B, A) \in \mathcal{R}1$. Hence, if (C,(A, B)) $\in \mathcal{D}$, then (A, B),$(B, A) \in \mathcal{R}$, i.e., Δ is preference symmetric.
2. If A undercuts B, then by Definition 32-3, there does not exist a (C,(A, B)) $\in \mathcal{D}$. Hence A S-defeats B for any $S \subset Args$.
3. Suppose ([ ],(X, Y)) $\in \mathcal{D}$. Since [ ] cannot conclude an ordering, then by Definition 32-3b, it must be that Y undercuts X. Hence $(Y, X) \in \mathcal{R}_1$, and so by Definition 32-3, $\neg\exists Z$ s.t. (Z,(Y, X)) $\in \mathcal{D}$.
4. By Definition 32-2c, B is incoherent implies ([ ],B) $\in \mathcal{R}3$. By Definition 27, [ ] does not conclusion-conclusion or undercut attack any argument, and so ([ ],B) cannot be in $\mathcal{R}1 \cup \mathcal{R}2$. By definition of an *EAF*, if (Z,(X, Y)), (Z',(Y,X)) $\in \mathcal{D}$), then (Z, Z'), $(Z', Z) \in \mathcal{R}$. Hence, suppose Z = [ ] and ([ ],(X, Y)) $\in \mathcal{D}$. By 3 there does not exist a (Z',(Y,X)) $\in \mathcal{D}$. Hence, ([ ],B) $\in \mathcal{R}$ **iff** B is incoherent.
5. By Definition 27, no argument can conclusion-conclusion or undercut [ ]. By 4, B cannot be [ ]. By 3 it cannot be that (B,[ ]) $\in \mathcal{R}$ on account of (B,(X,Y)), ([ ],(Y,X)) $\in \mathcal{D}$. Hence, there is no B such that (B,[ ]) $\in \mathcal{R}$.
6. It cannot be that A = B (i.e., A is incoherent) since this would imply (by assumption of B undercutting A) that A undercuts B. Hence, $(A, B) \in \mathcal{R}'$ in Definition 32-3, and since B undercuts A, ([ ],(A, B)) $\in \mathcal{D}$. Hence, for any S s.t. [ ] $\in S$, A does not defeat$_S$ B.
7. By Definition 32-2c, if A is incoherent then ([ ],A) $\in \mathcal{R}$. By Definition 32-3, no argument can *d*-attack ([ ],A), and so [ ] defeats$_S$ A for any S. By 5, no argument can attack and so defeat$_S$ [ ]. Hence, A cannot be acceptable w.r.t. any S.

Proposition 14 Let ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the *EAF* for a theory (S, D). If (C,(A, B)), (C',(B, A)) $\in \mathcal{D}$ then C and C' conclusion-conclusion attack.

Proof: By Lemma 11-3, $C \neq$ [ ], $C' \neq$ [ ]. By Definition 32-3, A does not undercut B, B does not undercut A, and (A, B), $(B, A) \in \mathcal{R}$ where A and B conclusion-conclusion attack. Without loss of generality, we can assume that they attack on a single pair of conclusions, such that R_B and R_A are the relevant sets of defeasible rules for B and A. By assumption of (C,(A,B)) $\in \mathcal{D}$, $\exists r \in R_A$ s.t. $\forall r' \in R_B$, C concludes $r \prec r'$ (denoted by $r \prec_C r'$). By assumption of (C',(B,A)) $\in \mathcal{D}$, $\exists r' \in R_B$ s.t. $\forall r \in R_A$, $r' \prec_{C'} r$. But then the latter can only be the case if for some $r \in R_A$, $r' \in R_B$, $r \prec_C r'$ and $r' \prec_{C'} r$. That is, C and C' conclusion-conclusion attack.

Lemmas 12 and 13 are required for the proof of the following Proposition 15.

Lemma 12 Let A' be a sub-argument of A, X an argument such that X conclusion-conclusion attacks both A and A', and A does not undercut X. Let B be an argument s.t. (B,(X,A)) $\in \mathcal{D}$. Then (B,(X,A')) $\in \mathcal{D}$

Proof: It cannot be the case that B = [ ], since B cannot conclude an ordering, and by Definition 32-3b, ([ ],(X,A)) $\in \mathcal{D}$ only when A undercuts X. Since A does not undercut X then no sub-argument A' of A undercuts X. It cannot be that X undercuts A, and so any sub-argument A' of A, since by Definition 32-3 we can not then have that (B,(X,A)) $\in \mathcal{D}$. Hence, let Ω = $\{(L_{A_1}, \overline{L_{X_1}}), \ldots, (L_{A_n}, \overline{L_{X_n}})\}$ be the set of pairs of conclusions on which A and X attack, (R_{A_1}, R_{X_1}),. . .,(R_{A_n}, R_{X_n}) the respective sets of relevant rules, where for $j = 1 \ldots n$, $R_{A_j} > R_{X_j}$, based on the ordering concluded by B. Since A' is a sub-argument of A, $\Omega' \subseteq \Omega$ is the set of pairs of conclusions on which A' and X attack. Hence (B,(X,A')) $\in \mathcal{D}$.

Lemma 13 Let Δ = $(Args, \mathcal{R}, \mathcal{D})$ be the *EAF* for (S,D), and let E be a conflict free subset of $Args$. If $A \in Args$ is acceptable w.r.t. E then all sub-arguments A' of A are acceptable w.r.t E.

Proof: Suppose for some sub-argument A' of A, $X \rightarrow^E A'$. By Lemma 11-7, A and so A' cannot be incoherent, and since ([],A') $\in \mathcal{R}$ iff A' is incoherent (by Lemma 11-4), then $X \neq$ [ ]. We show that there is a Z s.t. $Z \rightarrow^E X$ and a reinstatement set for $Z \rightarrow^E X$:
If $X \rightarrow^E A$ then by assumption of A acceptable w.r.t. E, there is a $Y \in E$, $Y \rightarrow^E X$ and there is a reinstatement set for $Y \rightarrow^E X$. Suppose $X \nrightarrow^E A$. Hence (by Lemma 11-2) X does not undercut A' or A, X conclusion-conclusion attacks A' and A, and $\exists B$ $\in E$ s.t. (B,(X,A)) $\in \mathcal{D}$. If A does not undercut X, then by Lemma 12, (B,(X,A')) $\in \mathcal{D}$, contradicting $X \rightarrow^E A'$. If A undercuts X, then $A \rightarrow^E X$ (by Lemma 11-2), and by Definition 32-3 there does not exist a $(C,(A,X)) \in \mathcal{D}$, and so we have the reinstatement set $\{A \rightarrow^E X\}$.

Proposition 15 Let Δ be the *EAF* for the ALP-DP theory (S,D). Then for $s \in$ $\{$grounded, complete, stable, preferred$\}$, if A is sceptically, respectively credulously, justified under the s semantics, then all sub-arguments of A are sceptically, respectively credulously, justified under the s semantics.

Proof: Follows immediately from Lemma 13.

Proposition 16 If (S,D) is priority-finitary then its *EAF* Δ = $(Args, \mathcal{R}, \mathcal{D})$ is finitary.

Proof: Since (S,D) is priority-finitary, any $Y \in Args$ is conclusion-conclusion or undercut attacked by a finite number of arguments, and so $\{X|(X,Y) \in \mathcal{R}\}$ is finite. Since, there is a finite number of rules deriving priorities, then the set of composite priority arguments is finite, and so $\{Z|(Z,(X,Y)) \in \mathcal{D}\}$ is finite. Hence, by Definition 12, Δ is finitary.

Proof of Proposition 17 below makes use of the following lemmas and notation. In what follows we will refer to [34]'s definition of arguments, attack, defeat, acceptability and an ALP-DP theory's characteristic function, as they are defined here in Definition 31.

Notation 3 S is said to be *closed under priority arguments* if $A_1 + \ldots + A_n$ is a composite priority argument in S iff $\{A_1, \ldots, A_n\} \subseteq S$, where for $i = 1 \ldots n$, A_i is a singleton priority argument.

Lemma 14 Given an ALP-DP theory (S, D), let $Args$ and $Args_{PS97}$[7] be the sets of arguments obtained by Definitions 25 and 31-1 respectively. Then $Args$ is closed under priority arguments, and $Args$ and $Args_{PS97}$ are *equivalent modulo composite priority arguments* (as defined in Definition 34).

Proof: Obvious, from Definitions 25 and 31-1.

Notation 4 In what follows we assume that: that:

- $S_{PS97}, T_{PS97} \ldots$ denote subsets of $Args_{PS97}$, where $Args_{PS97}$ denotes the arguments as defined by an ALP-DP theory in [34] (Definition 31-1 in this paper).
- Uppercase letters $S, T \ldots$ denote subsets of $Args$, where $Args$ is the arguments defined by an ALP-DP theory for instantiation of an *EAF* (i.e., $Args_{PS97}$ plus composite priority arguments as defined in Definition 25). Henceforth, we assume that for $X = Args, S, T \ldots$:
  X is closed under priority arguments, and X and X_{PS97} are equivalent modulo composite priority arguments.

Proposition 17 states that the grounded extension of an *EAF* Δ instantiated by an ALP-DP theory (S, D) (as defined in this paper) and the grounded extension of (S, D) as defined in [34], are equivalent modulo composite priority arguments (*emcpa*). We give an outline of the proof of the proposition:

1. Lemmas 15 and 17 are used in the proof of Lemma 18 which states that $A \in Args_{PS97}$ is acceptable w.r.t. an admissible X_{PS97} as defined in [34] iff A is strict-acceptable w.r.t. an admissible X as defined in this paper.

2. Corollary 5, following on from Lemma 19, states that the grounded extension of Δ is obtained by iterating the characteristic function F_{st} defined by strict acceptability, where the iteration starts with $\{[]\}$.

3. Lemma 21 shows that the grounded extension of (S, D), as defined in [34], is obtained by iterating the characteristic function G defined in [34], where the iteration starts with $\{[]\}$.

4. Given Corollary 5 and Lemma 21, Proposition 17 is proved by showing that starting with $\{[]\}$, the iterations of F_{st} and G yield sets that are *emcpa*. The proof is by induction on the iteration, whereby the base case is shown, and then the general case is shown by use of Lemma 18.

Lemma 15 Let $(Args, \mathcal{R}, \mathcal{D})$ be the *EAF* for an ALP-DP theory, $S \subseteq Args$ such that S is closed under priority arguments, and $[\ ] \in S$. Let $A, B \in Args$ such that A defeats$_S$ B, and A is a composite argument. Then, there exists a sub-argument A' of A such that A' is a singleton priority argument, and A' defeats$_S$ B.

Proof: Let $A = A_1 + \ldots + A_n$ where for $i = 1 \ldots n$, A_i is a singleton priority argument.

[7] PS97 refers to **P**rakken and **S**artor's paper [34] published in 19**97**.

If A defeats$_S$ B, then either:

• A undercuts B, in which case for some i, A_i undercuts B. By Lemma 11-2, A_i defeats$_S$ B.

• A does not undercut B, A conclusion-conclusion attacks B, and B does not undercut A (since if B did undercut A, then by Lemma 11-6, A would not defeat$_S$ B), and $\neg\exists C \in S$ s.t. (C,(A, B)) $\in \mathcal{D}$, i.e.:

$$\neg\exists C \in S \text{ s.t. } C \text{ concludes } \prec, \text{ and } B \text{ is preferred}_{\prec} \text{ to } A \qquad \textbf{(i)}$$

Let $\{A_k \ldots A_m\}$ be the singleton priority sub-arguments in A, such that for $j = k \ldots m$, A_j conclusion-conclusion attacks B. Without loss of generality we assume each such attack on a single pair $(L_j, \overline{L_j})$. Suppose that for $j = k \ldots m$, A_j does not defeat$_S$ B:

Since B does not undercut A, then for $j = k \ldots m$, B does not undercut A_j. Hence, for $j = k \ldots m$, $\exists C_j \in S$ such that C_j concludes $\prec_j$ and (by Definition 30), B is preferred$_{\prec_j}$ to A_j, where the relevant rule set for $\overline{L_j}$ is strictly greater than ($>$) the relevant rule set for L_j, and $>$ is defined on the basis of $\prec_j$ (as in Definition 29). Since S is closed under priority arguments, $C_k + \ldots + C_m$ is a composite priority argument in S that concludes $\prec = \bigcup_{j=k}^{m} \prec_j$, and so (by Definition 30) B is preferred$_\prec$ to A, contradicting **(i)**.

For Lemma 16, recall the definition of a priority argument C concluding an ordering $\prec$ (Definition 25-2), and [34]'s definition of an ordering $\prec_{T_{PS97}}$ concluded by singleton priority arguments in a set $T_{PS97} \subseteq Args_{PS97}$ (Definition 31-3).

Lemma 16 Let $S \subseteq Args$, $S_{PS97} \subseteq Args_{PS97}$. Then, $C \in S$ concludes $\prec$ iff there exists a $T_{PS97} \subseteq S_{PS97}$ such that $\prec_{T_{PS97}} = \prec$.

Proof: *Left to right half*: Suppose $C \in S$ concludes $\prec$, where $C = C_1 + \ldots C_n$. Since S and S_{PS97} are *emcpa*, then there is a $T_{PS97} = \{C_1, \ldots, C_n\} \subseteq S_{PS97}$, where $\prec_{T_{PS97}} = \prec$. *Right to left half*: Suppose $T_{PS97} \subseteq S_{PS97}$, and let $\{C_1, \ldots, C_n\}$ be the singleton priority arguments in T_{PS97} concluding $\prec_{T_{PS97}}$. Since S and S_{PS97} are *emcpa*, $\{C_1, \ldots, C_n\} \subseteq S$. Since S is closed under priority arguments, $C = C_1 + \ldots C_n \in S$, where C concludes $\prec = \prec_{T_{PS97}}$.

For Lemmas 17 and 18 recall [34]'s definition of defeat parameterised w.r.t. a set S_{PS97} of arguments (Definition 31-4).

Lemma 17 Let $(Args, \mathcal{R}, \mathcal{D})$ be the *EAF* for (S, D). Let $S \subseteq Args$, $S_{PS97} \subseteq Args_{PS97}$, such that [ ] $\in S$, [ ] $\in S_{PS97}$. $\forall A, B \in Args_{PS97}$, A S_{PS97}-defeats B iff A defeats$_S$ B.

Proof:

- Suppose B = [ ]:
  By Definition 31-2 and 31-4, B is not S_{PS97}-defeated by any argument.
  Given Lemma 11-5, B cannot be defeated$_S$ by any argument.

- Suppose A = [ ]:
  By Definitions 31-2 and 31-4, [ ] S_{PS97}-defeats B iff B is incoherent.
  By Lemma 11-4, ([ ],B) $\in \mathcal{R}$ iff B is incoherent. By Definition 32-3, there does not exist a (C,([ ],B)) $\in \mathcal{D}$. Hence, [ ] defeats$_S$ B iff B is incoherent.

- Suppose $A \neq$ [ ], $B \neq$ [ ]:

- If A undercuts B, then A S_{PS97}-defeats B (Definition 31-4b), and A defeats$_S$ B (Lemma 11-2).
- If A does not undercut B, A conclusion-conclusion attacks B and B does not undercut A, then:
  A S_{PS97}-defeats B iff B is not preferred$_{\prec_{S_{PS97}}}$ to A (Definition 31-4a).
  A defeats$_S$ B iff there is no $C \in S$ such that C concludes $\prec$ and B is preferred$_\prec$ to A (Definition 32-3a).
  Since preferred$_{\prec_{S_{PS97}}}$ and preferred$_\prec$ are defined in the same way, then given Lemma 16, B is not preferred$_{\prec_{S_{PS97}}}$ to A iff there is no $C \in S$ such that C concludes $\prec$ and B is preferred$_\prec$ to A. Hence, A S_{PS97}-defeats B iff A defeats$_S$ B.
- If A does not undercut B, A conclusion-conclusion attacks B and B undercuts A, then A does not S_{PS97}-defeats B (Definition 31-4a), and A does not defeat$_S$ B (Lemma 11-6).

Lemma 18 Let $\Delta = (Args, \mathcal{R}, \mathcal{D})$ be the *EAF* for (S, D). Let $S \subseteq Args$, $S_{PS97} \subseteq Args_{PS97}$, such that [ ] $\in S$, [ ] $\in S_{PS97}$, S is conflict free and S_{PS97} is conflict free. Suppose that every argument in S_{PS97} is ALP-DP acceptable w.r.t. S_{PS97}, and every argument in S is strict-acceptable w.r.t. S. Then, $\forall A \in Args_{PS97}$:

A is ALP-DP acceptable w.r.t. S_{PS97} iff A is strict-acceptable w.r.t. S.

Proof:

A is ALP-DP-acceptable w.r.t S_{PS97} iff $\forall B \in Args_{PS97}$ s.t. B S_{PS97}-defeats A, $\exists$ C $\in S_{PS97}$ s.t. C strictly S_{PS97}-defeats B

A is strict-acceptable w.r.t S iff $\forall B \in Args$ s.t. B defeats$_S$ A, $\exists$ $C \in S$ s.t. C strictly defeats$_S$ B.

Right to Left half: Assume $A \in Args_{PS97}$ is strict-acceptable w.r.t. S. Suppose a $B \in Args_{PS97}$ s.t. B S_{PS97}-defeats A. Since $Args$ and $Args_{PS97}$ are equivalent modulo composite priority arguments (*emcpa*), $B \in Args$. By Lemma 17, B defeats$_S$ A. By assumption of A being strict-acceptable w.r.t. S, $\exists C \in S$ s.t. C strictly defeats$_S$ B. There are two cases to consider:

1. $C \in Args_{PS97}$, $B \in Args_{PS97}$, and so by Lemma 17, C strictly S_{PS97}-defeats B. Since S and S_{PS97} are *emcpa*, $C \in S_{PS97}$. Hence, A is ALP-DP-acceptable w.r.t S_{PS97}.

2. $C \notin Args_{PS97}$. Hence, C is the composite priority argument $C_1 + \ldots + C_n$. By Lemma 15, there is a set $\{C_k, \ldots, C_m\}$ of singleton priority arguments that defeat$_S$ B. Since S and S_{PS97} are *emcpa*, $\{C_k, \ldots, C_m\} \subseteq S_{PS97}$. $B \in Args_{PS97}$, and so by Lemma 17, for $i = k \ldots m$, C_i S_{PS97}-defeats B. Suppose for some $i = k \ldots m$, C_i strictly S_{PS97}-defeats B. Then A is ALP-DP-acceptable w.r.t S_{PS97}. Suppose for $i = k \ldots m$, C_i does not strictly S_{PS97}-defeat B, i.e., B S_{PS97}-defeats C_i. But then by assumption of each C_i being ALP-DP-acceptable w.r.t. S_{PS97}, there must be an argument in S_{PS97} that strictly S_{PS97}-defeats B. Hence, A is ALP-DP-acceptable w.r.t S_{PS97}.

Left to Right half: Assume $A \in Args_{PS97}$ is ALP-DP acceptable w.r.t. S_{PS97}. Suppose some $B \in Args$ s.t. B defeats$_S$ A. We show that there is an argument in S that strictly defeats$_S$ B. There are two cases to consider:

1. $B \in Args_{PS97}$. By Lemma 17, B S_{PS97}-defeats A. By assumption of A being ALP-DP acceptable w.r.t. S_{PS97}, $\exists C \in S_{PS97}$ s.t. C strictly S_{PS97}-defeats B. Since S and S_{PS97} are *emcpa*, $C \in S$. Since $C, B \in Args_{PS97}$, then by Lemma 17, C strictly defeats$_S$ B. Hence, A is strict-acceptable w.r.t S.

2. $B \notin Args_{PS97}$. Hence, B is the composite priority argument $B_1 + \ldots + B_n$. By Lemma 15 there is a singleton B_i s.t. B_i defeats$_S$ A. Since $Args$ and $Args_{PS97}$ are *emcpa*, $B_i \in Args_{PS97}$. Hence, by Lemma 17, B_i S_{PS97}-defeats A, and since A is ALP-DP acceptable w.r.t. S, $\exists C \in S_{PS97}$ s.t. C strictly S_{PS97}-defeats B_i. Since S and S_{PS97} are *emcpa*, $C \in S$, and by Lemma 17, C strictly defeats$_S$ B_i. To show A is strict-acceptable w.r.t. S, we show that there must be an argument in S that strictly defeats$_S$ $B = B_1 + \ldots + B_n$:

   - Suppose B undercuts C. Then B defeats$_S$ C by Lemma 11-2. By assumption of $C \in S$ being strict-acceptable w.r.t. S, there is a $C' \in S$ s.t. C' strictly defeats$_S$ B.

   - Suppose B does not undercut C (and so B_i does not undercut C), and:

     - C undercuts B_i, and so C undercuts B. By Lemma 11-2, C defeats$_S$ B. If C and B do not conclusion-conclusion attack then C strictly defeats$_S$ B. Suppose C and B conclusion-conclusion attack. Since C undercuts B, then by Lemma 11-6, B cannot defeat$_S$ C, and so C strictly defeats$_S$ B.

     - C does not undercut B_i, C conclusion-conclusion attacks B_i, and so C conclusion-conclusion attacks B. Hence, (C, B), $(B, C) \in \mathcal{R}$, S is conflict free, and so by Lemma 3, either $C \rightarrow^S B$ or $B \rightarrow^S C$.
       Suppose $B \nrightarrow^S C$. Then C strictly defeats$_S$ B.
       Suppose $B \rightarrow^S C$. Then by assumption of C being strict acceptable w.r.t. S, there exists a $C' \in S$ s.t. C' strictly defeats$_S$ B.

Lemma 19 Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the *EAF* for a priority-finitary ALP-DP theory, and GE(Δ) the grounded extension of Δ. Let F be the characteristic function of Δ as defined in Definition 10. Let F^0_+ = $\{[\,]\}$, F^{i+1}_+ = $F(F^i_+)$. Then GE(Δ) = $\bigcup_{i=0}^{\infty}(F^i_+)$.

Proof: By Proposition 16, Δ is finitary. By Lemma 11-1, Δ is preference symmetric. Hence, by Proposition 12, GE(Δ) = $\bigcup_{i=0}^{\infty}(F^i)$, where F^0 = $\emptyset$, F^{i+1} = $F(F^i)$. Hence, GE(Δ) = $\bigcup_{i=0}^{\infty}(F^i_+)$ if the following holds:

$$\text{for } i = 0, 1, \ldots, F^i \subseteq F^i_+, \text{ and if } A \in F^i_+, A \notin F^i, \text{ then } A \in F^{i+1} \qquad \textbf{(1)}$$

Firstly, note that by Lemma 11-5, no argument can defeat$_S$ [ ], and so [ ] must be acceptable w.r.t. $\emptyset$. Hence, by Lemma 6:

$$\text{each } F^i_+ \text{ in the above defined sequence is conflict free.} \qquad \textbf{(2)}$$

Base Case:
i = 0: F^0 (= $\emptyset$) $\subseteq F^0_+$ (= $\{[\,]\}$), and by Lemma 11-5 no argument defeats$_\emptyset$ [ ], and so [ ] $\in F^1$.

i = 1: Trivially, F^0 and F^0_+ are conflict fee. $F^0 \subseteq F^0_+$ and F^1 = $F(F^0)$, F^1_+ = $F(F^0_+)$, and so by monotonicity of F (Proposition 11) $F^1 \subseteq F^1_+$.
F^1_+ and F^1 both contain all arguments that are not attacked by other arguments ($\{B | \forall A \in Args, (A, B) \notin \mathcal{R}\}$). F^1_+ - F^1 =
a) $\{B | \forall A \in Args$, if $(A, B) \in \mathcal{R}$ then ([ ],$(A, B)) \in \mathcal{D}\} \cup$
b) $\{B | \forall A \in Args$, if $A \rightarrow^{\{[\,]\}} B$ then [ ] $\rightarrow^{\{[\,]\}} A\}$. Note that by Lemma 11-5 A must be incoherent, and so by Definition 32-3 there is no $C \in Args$ s.t. (C,([ ],A))

$\in \mathcal{D}$, and so if [ ] $\rightarrow^{\{[]\}} A$ then [ ] $\rightarrow^{S} A$ for any $S \subseteq Args$, and $\{$[ ] $\rightarrow^{S} A\}$ is the reinstatement set for [ ] $\rightarrow^{S} A$.
Since [ ] $\in F^1$, then given a) and b), (F^1_+ - F^1) $\subseteq F^2$.

Inductive Hypothesis: **(1)** holds for $j < i$

General Case: $F^i \subseteq F^i_+$, and if $A \in F^i_+$, $A \notin F^i$, then $A \in F^{i+1}$:

By inductive hypothesis, $F^{i-1} \subseteq F^{i-1}_+$. By Proposition 6, F^{i-1} is conflict free. By **(2)**, F^{i-1}_+ is conflict free. Hence, by monotonicity of F, $F^i \subseteq F^i_+$.
By Proposition 6: a) $F^{i-1} \subseteq F^i$. By inductive hypothesis: b) $(F^{i-1}_+ - F^{i-1}) \subseteq F^i$. a) and b) imply: c) $F^{i-1}_+ \subseteq F^i$. Suppose $A \in F^i_+$ and $A \notin F^i$. We therefore have that: $A \in F(F^{i-1}_+)$; F^{i-1}_+ is conflict free by **(2)**; F^i is conflict free by Proposition 6; and so given c), then by monotonicity of F, $A \in (F(F^i) = F^{i+1})$.

Corollary 5 Let Δ = $(Args, \mathcal{R}, \mathcal{D})$ be the *EAF* for a priority-finitary ALP-DP theory, and GE(Δ) the grounded extension of Δ. Let F_{st} be the characteristic function of Δ based on strict acceptability (Definition 23). Let $F*^0_+$ = $\{[]\}$, $F*^{i+1}_+$ = $F_{st}(F*^i_+)$. Then, GE(Δ) = $\bigcup_{i=0}^{\infty}(F*^i_+)$.

Proof: Proof follows immediately from Lemma 19 and Corollary 3.

In the following lemmas 20 and 21 we refer to [34]'s definitions of an ALP-DP theory's arguments $Args$, the attack relation on $Args$, conflict free subsets of $Args$, T-defeat, ALP-DP-acceptability, an ALP-DP theory's characteristic function G, and its grounded extension (all of which are defined in this paper in Definition 31).

Lemma 20 Let (S, D) be an ALP-DP theory and T a conflict free subset of $Args$. Then $G(T)$ is conflict free.

Proof: Suppose $G(T)$ is not conflict free. Then $\exists A, B \in G(T)$ s.t. A attacks B, in which case A conclusion-conclusion or undercut attacks B, in which case one can straightforwardly show that A T-defeats B and/or B T-defeats A. By assumption of A and B acceptable w.r.t. S, there must exist arguments $C, D \in T$ s.t. C T-defeats D, in which case T is not conflict free. Contradiction.

Lemma 21 Let (S, D) be a priority-finitary ALP-DP theory and GE((S, D)) the grounded extension of (S, D) as defined as defined in [34] (defined in this paper in Definition 31-8). Then GE((S, D)) = $\bigcup_{i=0}^{\infty}(G^i_+)$, where G^0_+ = $\{[]\}$, G^{i+1}_+ = $G(G^i_+)$, and $G(G^i_+)$ = $\{A | A$ is ALP-DP-acceptable w.r.t $G^i_+\}$.

Proof: The lemma is shown by showing that for $i = 0, 1, \ldots$:
G^i and G^i_+ are conflict free, $G^i \subseteq G^i_+$, and if $A \in G^i_+$, $A \notin G^i$, then $A \in G^{i+1}$ **(1)**
Base case:

i = 0: Trivially, G^0 and G^0_+ are conflict fee, G^0 (= $\emptyset$) $\subseteq G^0_+$ (= $\{$[ ]$\}$), and by Definition 31-4 no argument $\emptyset$-defeats [ ], and so [ ] $\in G^1$.

i = 1: $G^0 \subseteq G^0_+$, and G^1 = $G(G^0)$, G^1_+ = $G(G^0_+)$, and since G^0 and G^0_+ are conflict fee, then by monotonicity of G (Proposition 5.5 in [34]) $G^1 \subseteq G^1_+$, and by Lemma 20 G^1 and G^1_+ are conflict free.
(G^1_+ - G^1) = $\{X | \forall Y s.t. Y \{$[ ]$\}$-defeats X, then [ ] $\{$[ ]$\}$-defeats $Y\}$. One can straightforwardly show that if [ ] $\{$[ ]$\}$-defeats Y then [ ] T-defeats Y for ny T. Hence, since [ ] $\in G^1$, (G^1_+ - G^1) $\subseteq G^2$.

Inductive Hypothesis: **(1)** holds for $j < i$

General Case: G^i and G^i_+ are conflict free, $G^i \subseteq G^i_+$, and if $A \in G^i_+$, $A \notin G^i$, then

$A \in G^{i+1}$:
By inductive hypothesis, G^{i-1} and G_+^{i-1} are conflict free, $G^{i-1} \subseteq G_+^{i-1}$, and so by monotonicity of G, $G^i \subseteq G_+^i$, and by Lemma 20 G^i and G_+^i are conflict free.
By monotonicity of G: a) $G^{i-1} \subseteq G^i$. By inductive hypothesis: b) $(G_+^{i-1} - G^{i-1}) \subseteq G^i$. a) and b) imply: c) $G_+^{i-1} \subseteq G^i$. Suppose $A \in G_+^i$ and $A \notin G^i$. Hence, $A \in G(G_+^{i-1})$, G_+^{i-1} is conflict free, G^i is conflict free, and so given c), then by monotonicity of G, $A \in (G(G^i) = G^{i+1})$.

Proposition 17 Let (S, D) be a priority-finitary ALP-DP theory, and let GE((S, D)) be the grounded extension of (S, D) as defined in Definition 31-8. Let Δ = ($Args$, $\mathcal{R}$, $\mathcal{D}$) be the theory's *EAF* and GE(Δ) the grounded extension of Δ. Then GE(Δ) and GE((S, D)) are equivalent modulo the composite priority arguments.

Proof: Let F_{st} be the characteristic function of Δ based on strict acceptability, and $F*_+^0$ = {[ ]}, $F*_+^{i+1}$ = $F_{st}(F*_+^i)$. Let G be the characteristic function of (S, D) and G_+^0 = {[ ]}, G_+^{i+1} = $G(G_+^i)$. Given Corollary 5 and Lemma 21, it is sufficient to show that for $\forall i$, $F*_+^i$ and G_+^i are equivalent modulo composite priority arguments (*emcpa*).
Base case:
G_+^1 = G({[ ]}) = $\{X | \forall Y s.t. Y$ {[ ]}-defeats X, then [ ] strictly {[ ]}-defeats $Y\} \cup$ the arguments not {[ ]}defeated by any argument, i.e.:

> a) the arguments B that are not undercut or conclusion-conclusion attacked by any argument, and;
> b) the arguments B that are not undercut by any argument, and for any A that conclusion-conclusion attacks B, B undercuts A.

$F*_+^1$ = F_{st}({[ ]}) = $\{X | \forall Y$ s.t. Y defeats$_{\{[]\}}$ X, then [ ] strictly defeats$_{\{[]\}}$ $Y\} \cup$ the arguments not defeated$_{\{[]\}}$ by any argument, i.e.:

> a) the arguments B that are not undercut or conclusion-conclusion attacked by any argument, i.e., $\{B | \forall A \in Args, (A, B) \notin \mathcal{R}\}$, and;
> b) the arguments B that are not undercut by any argument, and for any A that conclusion-conclusion attacks B ($(A, B) \in \mathcal{R}$), B undercuts A (and so ([ ],(A, B)) $\in \mathcal{D}$).

Note that:

> if $B1, \ldots, B_n$ are singleton priority arguments in $F*_+^1$ then B = $B_1 + \ldots + B_n$ is a composite priority argument in $F*_+^1$ **(1)**
>
> since: i) if any A conclusion-conclusion attacks B, then A conclusion-conclusion attacks some B_i, and so by b). B_i undercuts A, hence B undercuts A and so ([ ],(A, B)) $\in \mathcal{D}$; ii) $\forall Y$ s.t. Y defeats$_{\{[]\}}$ B, Y defeats$_{\{[]\}}$ some B_i, and so by assumption of $B_i \in F*_+^1$, [ ] strictly defeats$_{\{[]\}}$ Y.

Since $Args$ and $Args_{PS97}$ are *emcpa*, then given the above:
a) $G_+^1 \subseteq F*_+^1$
b) If $A \in F*_+^1$, $A \notin G_+^1$, then A must be a composite priority argument $A_1 + \ldots + A_n$ where $i > 1$, and for $i = 1 \ldots n$, A_i is a singleton priority argument.
c) By Lemma 13 and **(1)**, $A_1 + \ldots + A_n \in F*_+^1$ iff $\{A_1, \ldots, A_n\} \subseteq F*_+^1$, and so given a) and b):
d) $\{A_1, \ldots, A_n\} \subseteq G_+^1$.

a)-d) establish that $F*_+^1$ is closed under priority arguments, and $F*_+^1$ and G_+^1 are *emcpa*.

Inductive hypothesis: For $j < i$, $F{*}^j_+$ is closed under priority arguments, and $F{*}^j_+$ and G^j_+ are *emcpa*.

General Case: $F{*}^i_+$ and G^i_+ are *emcpa*.

We establish the following pre-conditions for applying Lemma 18:

- $F{*}^i_+ = F_{st}(F{*}^{i-1}_+)$ and $G^i_+ = G(G^{i-1}_+)$, where by inductive hypothesis $F{*}^{i-1}_+$ is closed under priority arguments, and $F{*}^{i-1}_+$ and G^{i-1}_+ are *emcpa*

- $[\,] \in F{*}^{i-1}_+$. By Lemma 6 and Corollary 3, $F{*}^{i-1}_+$ is conflict free and $F{*}^{i-1}_+ \subseteq F{*}^i_+$ (every argument in $F{*}^{i-1}$ is strict-acceptable w.r.t. $F{*}^{i-1}$).

- $[\,] \in G^{i-1}_+$, and by Lemma 20 and monotonicity of G, G^{i-1}_+ is conflict free, and every argument in G^{i-1}_+ is ALP-DP acceptable w.r.t. G^{i-1}_+

Given the above, then by Lemma 18:

$\forall A \in Args_{PS97}$, A is ALP-DP acceptable w.r.t. G^{i-1}_+ iff A is strict-acceptable w.r.t. $F{*}^{i-1}_+$ **(4)**

Suppose $A \in Args$, $A \notin Args_{PS97}$, and A is strict-acceptable w.r.t. $F{*}^{i-1}_+$. Then $A \in F{*}^i_+$ must be a composite priority argument $A_1 + \ldots + A_n$, where $i > 1$. By Lemma 13, $\{A_1, \ldots, A_n\} \subseteq F{*}^i_+$. Given **(4)** $\{A_1, \ldots, A_n\} \subseteq G^i_+$. Hence, $F{*}^i_+$ and G^i_+ are equivalent modulo composite priority arguments.

References

[1] L. Amgoud. A formal framework for handling conflicting desires. In *Proc. 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'2003)*, pages 552–563, Aalborg, Denmark, 2003.

[2] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning*, Volume 29 (2):125–169, 2002.

[3] 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.

[4] L. Amgoud, Y. Dimopolous, and P. Moraitis. A unified and general framework for argumentation-based negotiation. In *Proc. 6th International Joint Conference on Autonomous Agents and Multi-Agents Systems (AAMAS'2007)*, Honolulu, Hawai'i, 2007.

[5] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems (ICMAS-00)*, pages 31–38, 2000.

[6] ArguGRID. www.argugrid.org. July 2007 onwards.

[7] ASPIC. Argumentation services platform with integrated components (www.argumentation.org). March 2004 - September 2007.

[8] K. M. Atkinson, T. J. M. Bench-Capon, and P. McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.

[9] P. Baroni and M. Giacomin. Resolution-based argumentation semantics. In *Proc. 2nd International Conference on Computational Models of Argument*, pages 25–36, Toulouse, France, 2008.

[10] H. Barringer, D. M. Gabbay, and J. Woods. Temporal dynamics of support and attack networks: From argumentation to zoology. In *Mechanizing Mathematical Reasoning*, pages 59–98, 2005.

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

[12] T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171:10–15, 2007.

[13] 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.

[14] G. Brewka. Reasoning about priorities in default logic. In *Proceedings of the National Conference on Artificial Intelligence, AAAI'94*, pages 940–945, Seattle, Washington, 1994.

[15] C. Cayrol, S. Doutre, and J. Mengin. On Decision Problems Related to the Preferred Semantics for Argumentation Frameworks. *Journal of Logic and Computation*, 13(3):377–403, 2003.

[16] S. Coste-Marquis, C. Devred, and P. Marquis. Symmetric argumentation frameworks. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference (ECSQARU)*, pages 317–328, Barcelona, Spain, 2005.

[17] M. Dastani, G. Governatori, A. Rotolo, and L. W. N. van der Torre. Preferences of agents in defeasible logic. In *Australian Conference on Artificial Intelligence*, pages 695–704, Sydney, Australia, 2005.

[18] P. M. Dung. An argumentation semantics for logic programming with explicit negation. In *Proceedings of the Tenth Logic Programming Conference*, pages 616–630, Cambridge, MA, 1993.

[19] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[20] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.

[21] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In *Proc. 4th Biennial Conference of the Ontario Society for the Study of Argumentation*, 2001.

[22] J.Broersen, M.Dastani, J.Hulstijn, and L.W.N. van der Torre. Goal generation in the boid architecture. *Cognitive Science Quarterly Journal*, 2 (3-4):428–447, 2002.

[23] A. Kakas and P. Moraitis. Argumentative agent deliberation, roles and context. In *Computational Logic in Multi-Agent Systems: recent advances and future directions*, pages 35–48, 2002.

[24] A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proc. Second international joint conference on autonomous agents and multiagent systems*, pages 883–890, Melbourne, Australia, 2003.

[25] S. Modgil. Nested argumentation and its application to decision making over actions. In *Proc. Second Int. Workshop on Argumentation in Multi-Agent Systems*, pages 57–73, 2005.

[26] S. Modgil. Hierarchical argumentation. In *Proc. 10th European Conference on Logics in Artificial Intelligence*, pages 319–332, Liverpool, UK, 2006.

[27] S. Modgil. Value based argumentation in hierarchical argumentation frameworks. In *Proc. 1st Int. Conference on Computational Models of Argument*, pages 297–308, Liverpool, UK, 2006.

[28] S. Modgil. An abstract theory of argumentation that accommodates defeasible reasoning about preferences. In *Proc. 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 648–659, Tunisia, 2007.

[29] S. Modgil. An argumentation based semantics for agent reasoning. In *Proc. Workshop on Languages, methodologies and development tools for multi-agent systems*, pages 37–53, Durham, UK, 2007.

[30] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Itelligence (to appear)*, available at: www.dcs.kcl.ac.uk/staff/modgilsa/ProofTheoriesAlgorithms.pdf, 2009.

[31] S. Modgil and M. Luck. Argumentation based resolution of conflicts between desires and normative goals. In *Proc. Fifth Int. Workshop on Argumentation in Multi-Agent Systems (to appear)*, Portugal, 2008.

[32] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[33] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation*, 15:1009–1040, 2005.

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

[35] I. Rahwan and L. Amgoud. An argumentation based approach for practical reasoning. In *Proc. Fifth International Joint Conference on Autonomous agents and Multiagent systems*, pages 347–354, Hakodate, Japan, 2006.

[36] J.R. Searle. *Rationality in Action.* MIT Press, Cambridge, MA, 2001.

[37] G. A. W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proc. 7th European Workshop on Logic for Artificial Intelligence*, pages 239–253, 2000.