Two roles of preferences in argumentation frameworks

Leila Amgoud Srdjan Vesic

{amgoud,vesic}@irit.fr

Abstract. In this paper, we show that preferences intervene twice in argumentation frameworks: i) to compute standard solutions, and ii) to refine those solutions. The two roles are independent and obey to distinct postulates. After introducing and studying the postulates, we provide an example of a formal framework which models the two roles and verifies all the proposed postulates.

1 Introduction

An argumentation framework (AF) consists of a set of arguments and an attack relation among them. Arguments are evaluated using an acceptability semantics. This amounts to compute acceptable sets of arguments, called *extensions*. The attack relation is at the heart of all existing semantics. An attacker wins unless the attacked argument is defended by "good" arguments. Since [12], it has been argued that arguments may not have the same strength and some of them may be stronger or preferred to others. Consequently, several attempts have been made in the literature for taking into account preferences in argumentation frameworks (e.g. [2, 5]).

Besides, preferences play a key role in nonmonotonic reasoning [6]. They are used in order to narrow down the number of possible belief sets of a base theory. To say it differently, from a given base theory, a first set of *standard* solutions (belief sets) is computed, then a subset of those solutions (called *preferred* solutions) is chosen on the basis of available preferences. Thus, preferences refine the standard solutions.

In this paper, we show that preferences intervene twice in an argumentation framework. They are mandatory for: i) computing the standard solutions of an AF, and then ii) for narrowing the number of those solutions. The first role of preferences may not take into account *all* the available preferences. It focuses only on those which conflict with the attacks; such attacks are said *critical*. The idea is that an attack may fail if the attacked argument is stronger than its attacker. Ignoring this issue may lead to counter-intuitive standard solutions. The first role has largely been discussed in existing literature while the second one has only been pointed out recently in [7]. However, the difference between the two roles is still obscure. In this paper, we clarify the distinction between the two roles, and show that they are completely independent since none of them can be modeled by the other one. We propose postulates that should be satisfied by any preference-based argumentation framework. Some of them concern the first role while others concern the refinement role. Those postulates confirm again that the two roles are different. We propose a particular framework in which both roles are modeled. The properties of this framework are investigated.

The paper is structured as follows: We start by recalling Dung's AF, then we discuss informally the two roles of preferences. The two next sections propose postulates that

guide the definition of 'approaches' for each role. Then, we propose a particular framework which considers both roles. Before concluding, we compare our contribution with existing works. Due to lack of space, the proofs are not included in the paper.

2 Basics of Argumentation

The abstract argumentation framework proposed in [8] consists of a set of arguments and an attack relation between them.

Definition 1 (Argumentation framework) *An* argumentation framework *(AF) is a pair* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$*, where* $\mathcal{A}$ *is a set of arguments and* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is an attack relation. The notation* $a\mathcal{R}b$ *means that* a attacks b*.*

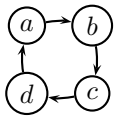
Different *acceptability semantics* for evaluating arguments have been proposed in the same paper [8]. Each semantics amounts to define sets of acceptable arguments, called *extensions*. The following definition recalls them.

Definition 2 (Stable semantics) *Let* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *be an AF and* $\mathcal{B} \subseteq \mathcal{A}$*.*

- $\mathcal{B}$ *is an* admissible *set iff it is conflict-free and defends all its elements.*
- $\mathcal{B}$ *is a* complete *extension iff it is conflict-free and contains all arguments it defends.*
- $\mathcal{B}$ *is a* preferred *extension iff it is a maximal (for set inclusion) admissible set.*
- $\mathcal{B}$ *is a* grounded *extension iff it is a minimal (for set inclusion) complete set.*
- $\mathcal{B}$ *is a* stable *extension iff it is a preferred set that attacks any element in* $\mathcal{A} \setminus \mathcal{B}$*.*

Let $\texttt{Ext}(\mathcal{F})$ *be the set of extensions of* $\mathcal{F}$ *under a given semantics.*

Example 1 *Let us consider the AF* $\mathcal{F}_1$ *depicted below. It has two stable extensions:* $\{a, c\}$ *and* $\{b, d\}$*.*



Definition 3 *Let* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *be an AF and* $\mathcal{E}_1, \ldots, \mathcal{E}_n$ *its extensions (under a given semantics). Let* $a \in \mathcal{A}$*.* a *is* skeptically accepted *iff there exists at least one extension and* $\forall i = 1, \ldots, n$*,* $a \in \mathcal{E}_i$*.* a *is* credulously accepted *iff* $\exists i = 1, \ldots, n$ *s.t.* $a \in \mathcal{E}_i$*.* a *is* rejected *iff* $\forall i = 1, \ldots, n$*,* $a \notin \mathcal{E}_i$*. Let* $\texttt{Status}(a, \mathcal{F})$ *be a function that returns the status of an argument* a *in* $\mathcal{F}$*.*

Example 1 (Cont): The four arguments a, b, c, d are credulously accepted in $\mathcal{F}_1$.

3 Preferences in Argumentation: Informal Discussion

In what follows, we assume that $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ is an arbitrary argumentation framework where $\mathcal{A}$ is *finite*. Let $\geq$ be a binary relation that expresses preferences between arguments of $\mathcal{A}$. For instance, an argument may be preferred to another if it is grounded on more certain information, or if it promotes a more important value. Throughout the paper, the relation $\geq \subseteq \mathcal{A} \times \mathcal{A}$ is assumed to be a *preorder* (i.e. *reflexive* and *transitive*). For arguments a and b, writing $a \geq b$ (or $(a, b) \in \geq$) means that a is at least as strong as b. The relation $>$ is the strict version of $\geq$. Indeed $a > b$ iff $a \geq b$ and not $(b \geq a)$.

Let us now analyze the role that preferences between arguments can play in an argumentation framework. We will discuss different critical examples.

Example 1 (Cont): If we assume that $a > b$ and $c > d$, then the stable extension $\{a, c\}$ is better than $\{b, d\}$ since each element of the latter is weaker than an element of the former. Thus, $\mathcal{F}_2$ would have only $\{a, c\}$ as its preferred stable extension.

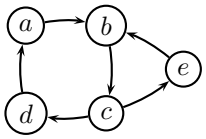
It is worth mentioning that preferences in Example 1 *refine* the results obtained in the standard case. Indeed, the set of preferred solutions is a *subset of the set of the standard ones*. Preferences play here exactly the role described in nonmonotonic logic formalisms. Let us now consider a different example.

Example 2 *Let $\mathcal{F}_2 = (\mathcal{A}, \mathcal{R})$ be s.t. $\mathcal{A} = \{a, b\}$ and $a\mathcal{R}b$. $\mathcal{F}_2$ has one stable extension: the set $\{a\}$. Now, if we assume that $b > a$, it is clear that the standard solution cannot be refined and $\{a\}$ is the preferred solution of the framework. What happened here is that the preferred argument is rejected when computing the standard solution. Thus, there is no way to apply the preference of b over a.*
However, is it intuitive to still consider $\{a\}$ as an extension of $\mathcal{F}_2$? The answer is certainly no as illustrated next. Assume that $\mathcal{F}_2$ is is built over a knowledge base $\mathcal{K} = \{x\}$ and a set of defeasible rules $\mathcal{D} = \{\Rightarrow y; y \Rightarrow \neg x\}$ as in ASPIC system [1]. Let $a :\Rightarrow y; y \Rightarrow \neg x$ and $b : x$. If the attack relation is the one which allows to undermine a premise of another argument, then a undermines b in its premise x while b does not undermine a since it has no premise. If now we assume that x is more certain than both $\Rightarrow y; y \Rightarrow \neg x$, then it is natural to keep b and to reject a. To put it differently, the preferred solution of $\mathcal{F}_2$ would be the extension $\{b\}$.

Contrarily to Example 1, the use of preferences in Example 2 completely modifies the original set of extensions. Consequently, the preferred solutions of a framework are not necessarily a subset of the standard ones. This is not surprising since preferences in this case are used in order to compute the standard solutions. Thus, $\{b\}$ is a standard solution. Preferred solutions refine the standard ones. In this example, $\{b\}$ is the only standard solution, thus it is also the unique preferred solution.

It is also worth mentioning that when preferences are used for computing the standard solutions of an argumentation framework, not all available preferences are exploited. Only those which conflict with the attacks, as in Example 2, are used. Consequently, the result which is returned may need to be refined as shown in the following example.

Example 3 *Let us consider the argumentation framework $\mathcal{F}_3$, which is depicted below.*



This framework has one stable extension which is $\{a, c\}$. Assume that $b > c$, $d > a$ and $b > e$. Note that only $b > e$ conflicts with the attack relation since $e\mathcal{R}b$. Thus, only this preference is taken into account for computing the two standard solutions $\{a, c\}$ and $\{b, d\}$. Consequently, the two remaining preferences can be used in order to refine the standard result and to prefer the extension $\{b, d\}$.

In sum, two roles of preferences are distinguished:

1. To weaken the *critical* attacks (i.e. the attacks which conflict with the preferences) in an AF, and thus to compute intuitive standard solutions.
2. To refine the standard solutions computed in the first role.

Example 2 shows that a refinement does not solve the problem of critical attacks and Example 3 shows that the first role is not sufficient and its results may need to be refined as the first role does not exploit all the available preferences.

4 Handling critical attacks

The aim of this section is to propose the basic postulates that any preference-based argumentation framework (PAF) should satisfy. We focus here on the use of preferences for computing the standard solutions, thus for modeling the first role of preferences.

Definition 4 (PAF) *A* preference-based argumentation framework *(PAF) is a tuple $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R}$ is an attack relation and $\geq$ is partial or total preorder on $\mathcal{A}$.*

Note that we do not show how arguments are evaluated in such a PAF. In fact, we do not focus on a particular approach, but we propose postulates that any approach should satisfy. Before presenting those postulates, let us first define formally critical attacks.

Definition 5 (Critical attack) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. An attack $(b, a) \in \mathcal{R}$ is* critical *iff $a > b$.*

The role of preferences which consists of handling critical attacks has already been identified in the literature, namely in [2, 4, 5, 10]. While all these approaches agree that a strong argument may be accepted if it is attacked by a weaker argument, they disagree on whether the weak attacker should be rejected or not. Let us say it differently, in Example 2, the works [2, 5, 10] return one stable extension which contains both the attacker and the attacked argument, that is the set $\{a, b\}$. This extension violates one of the basic requirements of acceptability semantics, the *conflict-freeness* of extensions. In

[4], the authors have argued that this is undesirable since the intuition behind an extension is that it encodes a 'coherent position'. This coherence is captured by the notion of conflict-freeness in acceptability semantics. That is why it is at the heart of all semantics. The authors have then proposed an alternative solution in which the argument a is rejected and the only stable extension of the framework $\mathcal{F}_2$ is $\{b\}$. In this paper, we argue that the extensions of an argumentation framework should be conflict-free, otherwise the whole theory of argumentation collapses. We propose four basic postulates that should be satisfied by any approach for preference-based argumentation that models the first role of preferences. The first postulates states that the extensions of a PAF should be conflict-free.

Postulate 1 (Conflict-freeness) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\texttt{Ext}(\mathcal{T})$ it set of extensions. Each extension $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$ should be conflict-free wrt $\mathcal{R}$.*

The second postulate says that when there are no critical attacks, then the output of the PAF should coincide with that of a system without preferences. The reason is that we suppose that a PAF is built over a well-founded basic system (i.e. the system constructed only from a pair $(\mathcal{A}, \mathcal{R})$).

Postulate 2 (Recovering existing semantics) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ its basic version. If there are no critical attacks in $\mathcal{T}$, then $\texttt{Ext}(\mathcal{T}) = \texttt{Ext}(\mathcal{F})$ where $\texttt{Ext}(\mathcal{F})$ is the set of the extensions of $\mathcal{F}$ under a given semantics.*

The third postulate shows how to privilege a strong argument over a weak attacker.

Postulate 3 (Critical attacks) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $a, b \in \mathcal{A}$. Let $\mathcal{E}_1, \mathcal{E}_2$ be two conflict-free (wrt $\mathcal{R}$) subsets of $\mathcal{A}$ s.t. $\mathcal{E}_1 = \mathcal{E} \cup \{a\}$ and $\mathcal{E}_2 = \mathcal{E} \cup \{b\}$. If $a\mathcal{R}b$ and $b > a$, then $\mathcal{E}_1 \notin \texttt{Ext}(\mathcal{T})$.*

The last postulate states that attacks should win when they are not critical.

Postulate 4 (Normal attacks) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $a, b \in \mathcal{A}$. Let $\mathcal{E}_1, \mathcal{E}_2$ be two conflict-free (wrt $\mathcal{R}$) subsets of $\mathcal{A}$ s.t. $\mathcal{E}_1 = \mathcal{E} \cup \{a\}$ and $\mathcal{E}_2 = \mathcal{E} \cup \{b\}$. If $a\mathcal{R}b$ and not($b\mathcal{R}a$) and not($b > a$), then $\mathcal{E}_2 \notin \texttt{Ext}(\mathcal{T})$.*

Works in [2, 5, 10], proceed by removing critical attacks from an argumentation graph and applying Dung's semantics on the remaining sub-graph. It is easy to show that when there are no critical attacks, the two graphs coincide.

Property 1. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF, $\geq \subseteq \mathcal{A} \times \mathcal{A}$, and $\mathcal{F}' = (\mathcal{A}, \mathcal{R}_r)$ be such that $\mathcal{R}_r = \mathcal{R} \setminus \{a\mathcal{R}b \text{ s.t. } b > a\}$. If $\nexists a, b \in \mathcal{A}$ s.t. $a\mathcal{R}b$ and $b > a$, then $\mathcal{R} = \mathcal{R}_r$.

It can be shown that that such an approach violates the conflict-freeness in some cases when the attack relation is not symmetric, and the third postulate (for example for admissible semantics), while it satisfies Postulates 2 and 4.

Proposition 1 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t. $\texttt{Ext}(\mathcal{T}) = \texttt{Ext}((\mathcal{A}, \mathcal{R}_r))$ where $\mathcal{R}_r = \mathcal{R} \setminus \{a\mathcal{R}b \text{ s.t. } b > a\}$. Then, $\mathcal{T}$ verifies Postulates 2 and 4.*

When the attack relation is symmetric, Postulates 1 and 3 are verified.

Proposition 2 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t. $\texttt{Ext}(\mathcal{T}) = \texttt{Ext}(\mathcal{F})$ where $\mathcal{F} = (\mathcal{A}, \mathcal{R}_r)$. If $\mathcal{R}$ is symmetric, then $\mathcal{T}$ verifies Postulates 1 and 3.*

This means that when the attack relation is symmetric, all the postulates are verified. However, the following example shows that the result may still need to be refined.

Example 4 *Let $\mathcal{A} = \{a, b, c, d\}$, $\mathcal{R} = \{(a, c), (c, a), (a, d), (d, a), (b, c), (c, b), (b, d), (d, b)\}$ and $a > c$, $b > d$. The extensions of this PAF are $\{a, b\}$ and $\{c, d\}$. However, $\{a, b\}$ is clearly preferred to $\{c, d\}$. Thus, the frameworks developed in [2, 5, 10] do not take into account the second role of preferences even when the attack relation is symmetric.*

In the recent paper ([4]) an approach has been proposed which verifies all postulates.

Proposition 3 *The class of PAFs defined in [4] verifies Postulates 1 - 4.*

5 Refining AFs by preferences

Until now, we have studied the first role of preferences. We have particularly shown that "some" preferences should be taken into account for computing the standard solutions of an argumentation framework. Examples 1 and 3 show that standard solutions may need to be narrowed down using the remaining preferences. What is worth noticing is that a refinement amounts to *compare* subsets of arguments. In Example 1, the so-called *democratic* relation, $\succeq_d$, is used for comparing the two sets $\{a, c\}$ and $\{b, d\}$:

$$\text{Let } \mathcal{E}, \mathcal{E}' \subseteq \mathcal{A}.\ \mathcal{E} \succeq_d \mathcal{E}' \text{ iff } \forall x' \in \mathcal{E}' \setminus \mathcal{E}, \exists x \in \mathcal{E} \setminus \mathcal{E}' \text{ s.t. } x > x'.$$

Relation $\succeq_d$ is not unique and different relations can be used as shown next.

Example 1 (Cont): Let us consider again $\mathcal{F}_1$ and assume that $a \approx b$ and $c > d$. According to relation $\succeq_d$, the two extensions $\{a, c\}$ and $\{b, d\}$ are incomparable. However, since $a \approx b$ and $c > d$, it is clear that one could prefer $\{a, c\}$ to $\{b, d\}$.

Let us now define the basic properties that such a relation should satisfy. The first property ensures that the refinement relation is a preorder, that is reflexive and transitive. Note that these are the basic properties of any preference relation.

Postulate 5 (Preorder) *Let $\mathcal{A}$ be a set of arguments. A refinement relation on $\mathcal{P}(\mathcal{A})$ is a preorder (reflexive and transitive).*

The second property ensures that the relation privileges sets that contain strong arguments (wrt).

Postulate 6 (Privileging strong arguments) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF, $a, b \in \mathcal{A}$ and $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{P}(\mathcal{A})$. If $\mathcal{E}_1 = \mathcal{E} \cup \{a\}$ and $\mathcal{E}_2 = \mathcal{E} \cup \{b\}$ and $a > b$, then $\mathcal{E}_1 \succeq \mathcal{E}_2$.*

Notation: The maximal elements of a set $\mathcal{S}$ wrt. a given relation $\succeq$ are defined as follows: $\texttt{Max}(\mathcal{S}, \succeq) = \{\mathcal{E} \subseteq \mathcal{S} \mid \nexists \mathcal{E}' \subseteq \mathcal{S} \text{ s.t. } \mathcal{E}' \succ \mathcal{E}\}$.

Property 2. The democratic relation verifies the two postulates 5 and 6.

6 A particular rich PAF

In this section, we propose a particular framework which models both roles of preferences and verifies all the postulates introduced in this paper. The framework follows two steps: at the first step, it computes the standard solutions by handling correctly the available critical attacks. These solutions are then refined using an appropriate refinement relation. In order to make the paper easy to read, we will call PAF the framework which computes the standard solutions and rich PAF the one which refines the results of the PAF.

Definition 6 (Rich PAFs) *A* rich PAF *is a tuple* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq, \succeq)$ *where* $\mathcal{A}$ *is a set of arguments,* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is an attack relation,* $\geq \subseteq \mathcal{A} \times \mathcal{A}$ *is a (partial or total) preorder and* $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ *is a relation which verifies Postulates 5 and 6. The extensions of* $\mathcal{T}$ *(under a given semantics) are elements of* $\texttt{Max}(\mathcal{S}, \succeq)$*, where* $\mathcal{S}$ *is the set of extensions (under the same semantics) of the PAF* $(\mathcal{A}, \mathcal{R}, \geq)$*.*

In what follows, we propose a new approach that handles correctly critical attacks (i.e. which satisfies the four postulates introduced in section 4). We exploit for that a simple result that is proved recently in [4]. In that paper, the authors have proposed a new approach for taking into account preferences and which prevents the shortcomings of existing ones, namely the problem of conflicting extensions. The basic idea is to integrate preferences in the definition of semantics. A refinement of stable semantics is defined as a dominance relation which compares sets of arguments. The best sets wrt that relation are the extensions of the PAF. In that paper, the authors have shown that all their extensions are conflict-free and Postulates 2, 3 and 4 as satisfied as well. They have also shown an important result for semantics that refine stable one with preferences. The result says that the extensions of their approach (i.e. the best sets wrt the dominance relation) are exactly the stable extensions of the basic argumentation framework in which each critical attack is inverted. In what follows, we show that this idea can be generalized to any acceptability semantics.

The idea of inverting the arrows of critical attacks in an argumentation graph allows to take into account the preference (between the two arguments involved in a critical attack) and in the same time the conflict between the two arguments of the attack is represented. The intuition behind this is that an attack between two arguments represents two things: i) an incoherence between the two arguments (in logic-based systems, it captures inconsistency between the supports of the two arguments), and ii) a kind of preference determined by the direction of the attack. Thus, in our approach, the direction of the arrow represents a real preference between arguments. Moreover, the conflict is kept between the two arguments. Dung's acceptability semantics are then applied on the modified graph. In our approach, standard solutions are computed by the following preference-based framework.

Definition 7 (PAF) *A* preference-based argumentation framework *(PAF) is a tuple* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ *where* $\mathcal{A}$ *is a set of arguments,* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is an attack relation and* $\geq$ *is a preorder on* $\mathcal{A}$*. The extensions of* $\mathcal{T}$ *under a given semantics are the extensions of the*

argumentation framework $(\mathcal{A}, \mathcal{R}_r)$*, called* repaired framework*, under the same semantics with:* $\mathcal{R}_r = \{(a,b)|(a,b) \in \mathcal{R}$ *and not* $(b > a)\} \cup \{(b,a)|(a,b) \in \mathcal{R}$ *and* $b > a\}$.

From Definition 7, it is clear that if a PAF has no critical attacks, then the repaired framework coincides with the basic one.

Property 3. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. If $\mathcal{T}$ has no critical attacks, then $\mathcal{R}_r = \mathcal{R}$.

This property shows also that when a PAF has no critical attacks, then preferences do not play any role in the evaluation process.

Our approach does not suffer from the drawback of existing ones. Indeed, it delivers conflict-free extensions of arguments. Thus, it satisfies Postulate 1.

Proposition 4 *Let* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ *be a PAF and* $\mathcal{E}_1, \ldots, \mathcal{E}_n$ *its extensions under a given semantics. For all* $i = 1, \ldots, n$*,* $\mathcal{E}_i$ *is conflict-free wrt.* $\mathcal{R}$*.*

The next result confirms that our approach is *well-founded* in the sense that acceptable arguments are defended by "good" arguments. Moreover, it verifies the orderings between the attack relation and the preference relation, meaning that it verifies Postulates 3 and 4.

Proposition 5 *Let* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ *be a PAF.*

- *For each admissible set* $\mathcal{E}$ *of* $\mathcal{T}$*, it holds that* $(\forall x \in \mathcal{E})$ $(\forall x' \notin \mathcal{E})$ *if* $(x'\mathcal{R}x$ *and not* $(x > x'))$ *or* $(x\mathcal{R}x'$ *and* $x' > x)$ *then* $(\exists y \in \mathcal{E})$ *s.t.* $(y\mathcal{R}x'$ *and not* $(x' > y))$ *or* $(x'\mathcal{R}y$ *and* $y > x')$*.*
- *For each stable extension* $\mathcal{E}$ *of* $\mathcal{T}$*, it holds that* $(\forall x' \notin \mathcal{E})$ $(\exists x \in \mathcal{E})$ *s.t.* $(x\mathcal{R}x'$ *and not* $(x' > x))$ *or* $(x'\mathcal{R}x$ *and* $x > x')$*.*

The fact of inverting the arrows of critical attacks in an argumentation graph does not affect the status of arguments that are not related to the arguments of those attacks. This means that our approach has no side effects. Before presenting the formal result, let us first give a useful definition.

Definition 8 *Let* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *be an AF and* $a, b \in \mathcal{A}$*. The arguments* a *and* b *are* related *in* $\mathcal{F}$ *iff there is exists a finite sequence* $a_1, \ldots, a_n$ *of arguments such that* $a_1 = a$*,* $a_n = b$ *and for all* $i = 1, \ldots, n-1$*, either* $(a_i, a_{i+1}) \in \mathcal{R}$ *or* $(a_{i+1}, a_i) \in \mathcal{R}$*.*

Proposition 6 *Let* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ *be a PAF. For all* $a \in \mathcal{A}$ *s.t.* $\nexists b, c \in \mathcal{A}$ *s.t.* $(b,c) \in \mathcal{R}$ *is a critical attack and* a *is related with* b*, it holds that:*

- $\texttt{Status}(a, (\mathcal{A}, \mathcal{R}))$ = $\texttt{Status}(a, (\mathcal{A}, \mathcal{R}_r))$ *(under preferred and grounded semantics).*
- *If* $(\mathcal{A}, \mathcal{R})$ *and* $(\mathcal{A}, \mathcal{R}_r)$ *both have at least one stable extension, then* $\texttt{Status}(a, (\mathcal{A}, \mathcal{R}))$ = $\texttt{Status}(a, (\mathcal{A}, \mathcal{R}_r))$ *(under this semantics).*

Our approach privileges the strongest arguments. Indeed, we show that these arguments are skeptically accepted when they are not conflicting. If such a strong argument is not skeptically accepted, then it is for sure attacked (wrt. $\mathcal{R}$) by another strongest argument. Before presenting the formal result, let us define the strongest arguments (or the top elements) wrt. a relation $\geq$.

Definition 9 (Maximal elements) *Let $\mathcal{O}$ be a set of objects and $\geq \subseteq \mathcal{O} \times \mathcal{O}$ is a (partial or total) preorder. The* maximal elements *of $\mathcal{O}$ wrt. $\geq$ are* $\texttt{Max}(\mathcal{O}, \geq) = \{o \in \mathcal{O} \mid \nexists o' \in \mathcal{O}$ *s.t.* $o' > o\}$.

Property 4. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t $\geq$ is complete[1].

- If $\texttt{Max}(\mathcal{A}, \geq)$ is conflict-free (wrt. $\mathcal{R}$), then $\forall a \in \texttt{Max}(\mathcal{A}, \geq)$:
  - a is skeptically accepted in $\mathcal{T}$ wrt. preferred and grounded semantics.
  - if $\mathcal{T}$ has at least one stable extension, then a is skeptically accepted wrt. stable semantics.
- If a is not skeptically accepted (under preferred or grounded semantics), or there exists at least one stable extension and a is not skeptically accepted, then $\exists b \in \texttt{Max}(\mathcal{A}, \geq)$ s.t. $(b, a) \in \mathcal{R}$.

The following result shows that when the preference relation $\geq$ is a linear order (i.e. reflexive, antisymmetric, transitive and complete), then the corresponding PAF has a unique stable/preferred extension. Moreover, this extension is computed in $\mathcal{O}(n^2)$ time where $|\mathcal{A}| = n$. It is clear that in this case, there is no need to refine the result.

Proposition 7 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t. $\mathcal{R}$ is irreflexive and $\geq$ is a linear order.*

- *$\mathcal{T}$ has exactly one stable extension.*
- *Stable, preferred and grounded extensions of $\mathcal{T}$ coincide.*
- *If $|\mathcal{A}| = n$, then this extension is computed in $\mathcal{O}(n^2)$ time.*

Let us now see what happens in case the attack relation is symmetric. The following result shows that our approach returns the same results as the approach developed in [2, 5]. This means that inverting the arrows or removing them will lead to the same result.

Property 5. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF where $\mathcal{R}$ is symmetric. Extensions of $\mathcal{T}$ coincide with extensions of $(\mathcal{A}, \mathcal{R}')$ (under the same semantics) where $\mathcal{R}' = \{(a, b) | (a, b) \in \mathcal{R}$ and$\neg(b > a)\}$.

We can also show that when the attack relation is symmetric, the extensions of a PAF are a subset of those of its basic framework. This means that preferences filter the extensions. However, the result is not optimal since it may need to be refined again as shown in Example 4.

Proposition 8 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF where $\mathcal{R}$ is symmetric. If $\mathcal{E} \subseteq \mathcal{A}$ is a preferred (stable) extension of system $\mathcal{T}$ then $\mathcal{E}$ is a preferred (stable) extension of $(\mathcal{A}, \mathcal{R})$.*

Recall that this result is not true in case the attack relation is not symmetric as shown in Example 2.

The following result characterizes the extensions of $(\mathcal{A}, \mathcal{R})$ that are discarded in a PAF when $\mathcal{R}$ is symmetric. The idea is that an extension is discarded iff some argument outside it is strictly preferred to any arguments of that extension with which it conflicts.

[1] A relation $\geq$ on a set $\mathcal{A}$ is complete iff for all $a, b \in \mathcal{A}$, $a \geq b$ or $b \geq a$.

Property 6. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t. $\mathcal{R}$ is symmetric, and $\mathcal{E} \subseteq \mathcal{A}$. $\mathcal{E}$ is a stable extension of $(\mathcal{A}, \mathcal{R})$ but not of $\mathcal{T}$ iff $\exists x' \notin \mathcal{E}$ s.t. $\forall x \in \mathcal{E}$, if $x\mathcal{R}x'$, then $x' > x$.

When the attack relation is symmetric and irreflexive, the corresponding PAF is *coherent* (i.e. its preferred and stable extensions coincide) and it has at least one stable extension.

Proposition 9 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. If $\mathcal{R}$ is symmetric and irreflexive, then:*

- *$\mathcal{T}$ is coherent.*
- *$\mathcal{T}$ has at least one stable extension.*

Until now, we have proposed a particular framework for handling the first role of preferences. From now on, we will use the democratic relation for refining the results of this framework. Recall that this relation verifies the two postulates 5 and 6.

We will now show that when the preference relation $\geq$ is a linear order, then the democratic relation does not change the output of the underlying PAF.

Property 7. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq, \succeq)$ be a rich PAF and $\mathcal{S}$ be the set of extensions (under a given semantics) of the repaired framework $(\mathcal{A}, \mathcal{R}_r)$. If $\mathcal{R}$ is irreflexive and $\geq$ is a linear order, then $\texttt{Max}(\mathcal{S}, \succeq) = \mathcal{S}$ holds for stable, preferred, grounded and complete semantics.

It is also easy to show that when a rich PAF has no critical attacks, then its extensions are a subset of the extensions of its basic version (i.e. without preferences).

Property 8. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq, \succeq)$ be a rich PAF s.t. $\mathcal{R}$ has no critical attacks. Preferred (stable) extensions of $\mathcal{T}$ are exactly the elements of $\texttt{Max}(\mathcal{S}, \succeq)$ where $\mathcal{S}$ is the set of all preferred (stable) extensions of the AF $(\mathcal{A}, \mathcal{R})$.

Example 1 (Cont): Let us use the democratic relation $\succeq_d$. In $\mathcal{F}_1$, there is no critical attacks ($\mathcal{R}_r = \mathcal{R}$). The extensions of the rich PAF are $\texttt{Max}(\{\{a, c\}, \{b, d\}\}, \succeq_d) = \{\{a, c\}\}$. Thus, $\{a, c\}$ is the unique stable extension.

Example 2 (Cont): The repaired framework of $\mathcal{F}_2$ is $(\{a, b\}, \mathcal{R}_r)$ where $b\mathcal{R}a$. Thus, the PAF has one stable extension $\{b\}$ which is the only extension of the rich PAF: $\texttt{Max}(\{\{b\}\}, \succeq_d) = \{\{b\}\}$.

Example 3 (Cont): Recall that the repaired framework of $\mathcal{F}_3$ has two stable extensions: $\{a, c\}$ and $\{b, d\}$. Moreover, $\texttt{Max}(\{\{a, c\}, \{b, d\}\}, \succeq_d) = \{\{a, c\}\}$. Thus, $\{a, c\}$ is the unique stable extension of the rich PAF that uses the democratic relation.

7 Related work

Introducing preferences in argumentation frameworks goes back to the paper by Simari and Loui in [12]. In that work, the authors have defined an AF in which arguments are built from a propositional knowledge base. The arguments grounded on specific information are considered as stronger than the ones built from more general information.

This preference is used to solve dilemmas between any pair of conflicting arguments. Thus, it is used for handling critical attacks. The idea of this paper has been generalized in [2] then in [5] to any AF and to any preference relation. Unfortunately, the approach followed in [2, 5] delivers correct results only when the attack relation is symmetric. When the attack relation is not symmetric, the approach suffers from two main drawbacks: the first is that it may return conflicting extensions as shown in Example 1 since it may put two conflicting arguments in the same extension. One of these arguments is clearly undesirable. The second drawback is a consequence of the first one. Indeed, since an undesirable argument may be accepted, then all the arguments that are defended by this argument are accepted as well at the detriment of good ones. Let us illustrate this issue on the following example.

Example 5 *Let us consider the AF depicted in the left side of the figure below.*

Assume that $a > b$. The approach in [2, 5] gets the framework depicted in the right side of the same figure. Its grounded extension is the set $\{a, b, d\}$. This result is incorrect for two reasons: The first one is that the two arguments a and b cannot be both accepted. The second reason is that the argument b (which should be rejected) defends d against c, leading thus to an undesirable result. Indeed, d is defended by a "bad" argument! It is easy to check that our approach returns $\{a, c\}$ as the grounded extension and rejects the two other arguments: i.e. b and c.

Our approach overcomes the limits of the one proposed in [2, 5]. Moreover, it is more general since it models even the second role of preference (i.e. the refinement).
Recently, in [3], the authors have pointed out the first limit of the approach followed in [2, 5], namely the violation of conflict-freeness. They have proposed a new approach for handling critical attacks where preferences are introduced at a semantics level. As shown in this paper, the approach developed in [3] satisfies the four rationality postulates. However, it completely neglects the second role of preferences, i.e. refinement. Another work which handles correctly the problem of critical attacks is that proposed in [11]. In that paper, Prakken has proposed a logic-based instantiation of Dung's framework in which three kinds of attacks are considered: rebuttal, assumption attack and undercut. For each relation, the author has found a way to avoid the problem of critical attack and ensured conflict-free extensions. We think that our work is more general since we solved the problem at an abstract level. This avoid the user who wants to use another attack relation to look for new ways to avoid conflicting extensions. Moreover, our approach is axiomatic, meaning that it is well founded. It is also worth mentioning that in [11], the second role of preferences is neglected. To the best of our knowledge, the only work on refinement is that appeared in [7]. The authors have proposed a particular refinement relation in case of stable semantics. In this sense, our work is more general since it accepts any refinement relation. Moreover, there is no restriction to particular semantics. Finally, we would like to mention the work done in [9]. In this paper, the author made a survey of the critics presented in [3, 7] against existing approaches for PAFs. The author concluded that one should use a symmetric attack relation in order

to avoid the problem of conflicting extensions and then to refine the result with the preference relation already mentioned in [7]. The first suggestion is certainly not realistic, especially in light of new results in the literature stating that symmetric relations should be avoided in logic-based argumentation systems.

8 Conclusion

This paper has studied deeply the difference between the two roles that preferences may play in an AF. We have shown that preferences intervene both for computing what is called standard solutions in nonmonotonic reasoning formalisms and for refining that result, and choosing a subset of those solutions. We have shown that the two roles are completely independent and should be taken into account in two steps. Main postulates that any approach modeling each role have been proposed. Finally, we have developed a particular framework that considers both roles. The framework satisfies the proposed postulates and its properties show that it is well-founded.

References

1. L. Amgoud, M. Caminada, C. Cayrol, , M. Lagasquie, and H. Prakken. Towards a consensual formal model: inference part. Technical report, Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making. ASPIC project, 2004.
2. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Int.*, 34:197–216, 2002.
3. L. Amgoud and S. Vesic. Repairing preference-based argumentation systems. In *IJCAI'09*, pages 665–670, 2009.
4. L. Amgoud and S. Vesic. Generalizing stable semantics by preferences. In *Proceedings of COMMA'10*, pages 39–50, 2010.
5. T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. of Logic and Computation*, 13(3):429–448, 2003.
6. G. Brewka, I. Niemela, and M. Truszczynski. Preferences and nonmonotonic reasoning. *AI Magazine*, pages 69–78, 2008.
7. Y. Dimopoulos, P. Moraitis, and L. Amgoud. Extending argumentation to make good decisions. In *Proceedings of ADT'09*, pages 225–236. LNCS 5783, 2009.
8. P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AIJ*, 77:321–357, 1995.
9. S. Kaci. Refined preference-based argumentation framworks. In *Proceedings of COMMA'10*, pages 299–310, 2010.
10. S. Modgil. Reasoning about preferences in argumentation frameworks. *AIJ*, 173(9-10):901–934, 2009.
11. H. Prakken. An abstract framework for argumentation with structured arguments. *Journal of Argument and Computation*, page In Press, 2010.
12. G. Simari and R. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence Journal*, 53:125–157, 1992.