Generalizing stable semantics by preferences

Leila AMGOUD [a,1] and Srdjan VESIC [a]
[a] *Institut de Recherche en Informatique de Toulouse*

Abstract.
Different proposals have been made in the literature for refining Dung's argumentation framework by preferences between arguments. The idea is to ignore an attack if the attacked argument is stronger than its attacker. Acceptability semantics are then applied on the remaining attacks. Unfortunately, these proposals may return some unintended results, in particular, when the attack relation is asymmetric.
In this paper, we propose a new approach in which preferences are taken into account at the semantics level. In case preferences are not available or do not conflict with the attacks, the extensions of the new semantics coincide with those of the basic ones. Besides, in our approach, the extensions (under a given semantics) are the maximal elements of a *dominance relation* on the powerset of the set of arguments. Throughout the paper, we focus on stable semantics. We provide a full characterization of its dominance relations; and we refine it with preferences.

1. Introduction

Argumentation is a reasoning model based on the construction and the evaluation of arguments. An argument gives a reason to believe a statement, to perform an action, etc.

The most abstract argumentation framework in the literature has been proposed in [8]. It consists of a set of *arguments* and a binary relation that captures *attacks* among them. Different *acceptability semantics* have been proposed in the same paper. A semantics amounts to define sets of acceptable arguments, called *extensions*. In this framework, arguments are assumed to have all the same strength. Besides, in [5,7,11], it has been argued that some arguments may be stronger than others. In [2], a first *abstract* preference-based argumentation framework (PAF) has been proposed. It takes as input a set of arguments, an attack relation, and a preference relation between arguments which is abstract and can thus be instantiated in different ways. This proposal has been generalized in [10] in order to reason even about preferences. Thus, arguments may support preferences about arguments. The last extension has been proposed in [4]. It assumes that each argument promotes a value, and a preference between two arguments comes from the importance of the respective values that are promoted by the two arguments. Whatever the source of the preference relation is, the idea is to ignore an attack if the attacked argument is stronger than its attacker. Dung's semantics are then applied on the remaining attacks. Unfortunately, these proposals may return some unintended results, in

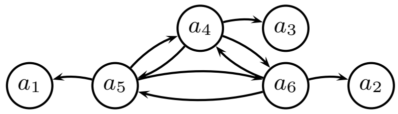
[1]Corresponding Author: amgoud@irit.fr.

particular, when the attack relation is not symmetric. Besides, in [1], it has been shown that the attack relation should not be symmetric because otherwise the corresponding argumentation framework violates the postulate on consistency [6].

Example 1 *Let* $\Sigma = \{x, \neg y, x \rightarrow y\}$ *be a propositional knowledge base s.t.* x *is* more certain *than the two other formulas. The following arguments*[2] *are built from this base:*

$a_1 :< \{x\}, x >$	$a_2 :< \{\neg y\}, \neg y >$
$a_3 :< \{x \rightarrow y\}, x \rightarrow y >$	$a_4 :< \{x, \neg y\}, x \wedge \neg y >$
$a_5 :< \{\neg y, x \rightarrow y\}, \neg x >$	$a_6 :< \{x, x \rightarrow y\}, y >$

The figure below depicts the attacks wrt "assumption attack"[3] *[9].*



Finally, assume that arguments are compared using the weakest link principle[4] *[5]. According to this relation, the argument* a_1 *is strictly preferred to the others, which are themselves equally preferred. The classical approaches of PAFs remove the attack from* a_5 *to* a_1 *and get* $\{a_1, a_2, a_3, a_5\}$ *as a stable extension. Note that this extension, which intends to support a* coherent point of view*, is conflicting since it contains both* a_1 *and* a_5 *and supports thus* x *and* $\neg x$.

In this paper, we propose a new approach for PAFs in which preferences are taken into account at the semantics level. The idea is that, instead of modifying the inputs of Dung's framework, we extend the semantics with preferences. In case these preferences are not available or do not conflict with the attacks, the extensions of the new semantics coincide with those of the basic ones. Besides, in our approach, the extensions (under a given semantics) are the maximal elements of a *dominance relation* on the powerset of the set of arguments. A dominance relation encodes thus an acceptability semantics in our case. Contrarily to existing semantics which partition the powerset of arguments into two subsets: the extensions and the non-extensions, our approach provides more information since it compares all susbsets of arguments. Another novelty of our approach is that it defines a semantics through a set of postulates. The postulates describe the desirable properties of a dominance relation. In this paper, we focus only on stable semantics. We provide a full characterization of its dominance relations; and we refine it with preferences. A representation theorem is given; it describes the extensions of the new semantics, called *pref-stable*.

The paper is organized as follows: The next section recalls Dung's framework. Then, we propose our new approach for PAFs. Next, we characterize the dominance relations that encode stable semantics. Then, we show how to refine stable semantics with preferences. The last section is devoted to some concluding remarks and future work.

[2]An *argument* is a pair $< H, h >$ where H is its *support* and h its *conclusion*. H is a minimal subset of Σ that is consistent and infers classically h.

[3]An argument a *attacks* b iff the conclusion of a is the contrary of a formula in the support of b.

[4]An argument a is *preferred* to an argument b if the least certain formula in the support of a is more certain than the least certain formula in the support of b.

2. Basic argumentation framework

In the seminal paper [8], an *argumentation framework* (AF) is a pair $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A}$ is a set of *arguments* and $\mathcal{R}$ is an *attack relation* between arguments ($\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$). The notation $(a, b) \in \mathcal{R}$ or $a\mathcal{R}b$ means that the argument a attacks the argument b. Different *acceptability semantics* for evaluating arguments have been proposed in the same paper. Each semantics amounts to define sets of acceptable arguments, called *extensions*. For the purpose of our paper, we only need to recall *stable* semantics.

Definition 1 (Conflict-free, Stable semantics) *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{B} \subseteq \mathcal{A}$.*

- *$\mathcal{B}$ is* conflict-free *iff $\nexists\, a, b \in \mathcal{B}$ s.t. $a\mathcal{R}b$.*
- *$\mathcal{B}$ is a* stable extension *iff it is conflict-free and attacks any argument in $\mathcal{A} \setminus \mathcal{B}$.*

$\mathtt{Ext}(\mathcal{F})$ *denotes the set of stable extensions of $\mathcal{F}$.*

Note that some argumentation frameworks may not have stable extensions.

3. A new approach for PAFs

A *preference-based argumentation framework* (PAF) takes as input three elements: a set $\mathcal{A}$ of arguments, a binary relation $\mathcal{R}$ capturing attacks between arguments, and a (partial or total) preorder[5] $\geq$ on the set $\mathcal{A}$. This latter encodes differences in strengths of arguments. The expression $(a, b) \in \geq$ or $a \geq b$ means that the argument a is at least as strong as b. The symbol $>$ denotes the strict relation associated with $\geq$. Indeed, $a > b$ iff $a \geq b$ and not ($b \geq a$).

Definition 2 (PAF) *A* 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 on $\mathcal{A}$, and $\geq$ is a (partial or total) preorder on $\mathcal{A}$.*

The new approach amounts to define *new acceptability semantics* that take into account the preference relation between arguments. A semantics is defined by a *dominance relation*, denoted by $\succeq$, on the powerset $\mathcal{P}(\mathcal{A})$ of the set of arguments. We say also that a dominance relation $\succeq$ *encodes* a semantics. For $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$, writing $(\mathcal{E}, \mathcal{E}') \in \succeq$ (or equivalently $\mathcal{E} \succeq \mathcal{E}'$) means that the set $\mathcal{E}$ is at least as good as the set $\mathcal{E}'$. The relation $\succ$ is the strict version of $\succeq$, that is for $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$, $\mathcal{E} \succ \mathcal{E}'$ iff $\mathcal{E} \succeq \mathcal{E}'$ and not ($\mathcal{E}' \succeq \mathcal{E}$).

Like the basic semantics of Dung, the new semantics computes extensions of arguments. These latter are the maximal elements of the dominance relation $\succeq$ that encodes the semantics. The notion of maximality is defined as follows.

Definition 3 (Maximal elements) *Let $\mathcal{E} \in \mathcal{P}(\mathcal{A})$ and $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$. $\mathcal{E}$ is* maximal *wrt $\succeq$ iff $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A})$, $\mathcal{E} \succeq \mathcal{E}'$.*

As we will see in the next sections, not any relation $\succeq$ can be used for evaluating arguments in a PAF. An appropriate relation should, for instance, ensure the conflict-freeness of its maximal elements. Recall that this property is at the heart of all Dung's semantics, as it avoids inconsistent conclusions.

[5] A binary relation is a *preorder* iff it is *reflexive* and *transitive*.

Definition 4 (Extensions of a PAF) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF, and $\mathcal{E} \in \mathcal{P}(\mathcal{A})$. The set $\mathcal{E}$ is an* extension *of $\mathcal{T}$ under the dominance relation $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ iff $\mathcal{E}$ is a maximal element of $\succeq$.*
Let $\texttt{Ext}_{\succeq}(\mathcal{T})$ denote the set of extensions of $\mathcal{T}$ wrt $\succeq$.

Notation: Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. $\mathcal{CF}(\mathcal{T})$ denotes the conflict-free (wrt $\mathcal{R}$) sets of arguments. At some places, we abuse notation and use $\mathcal{CF}(\mathcal{F})$ to denote the conflict-free sets of arguments of a basic framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$.

Assumptions: Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. Throughout the paper, we assume that:

1. The set $\mathcal{A}$ is finite.
2. $\mathcal{T}$ does not contain self-attacking arguments.

In the remainder of the paper, we will propose a new acceptability semantics, called *Pref-stable*. This semantics generalizes stable semantics with preferences between arguments. In case preferences are not available or are not conflicting with the attacks, the extensions of the two semantics coincide.

4. Stable semantics as a dominance relation

In the previous section, we have shown that our new semantics are defined as dominance relations on the power set of the set of arguments. The new semantics should recover the basic semantics of Dung in some cases. Before showing how to extend stable semantics with preferences, it is important to encode this semantics in the new setting, i.e. to define it as a dominance relation on the power set of the set of arguments. The following theorem characterizes the dominance relations that encode stable semantics.

Theorem 1 *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$. Then, $\forall \mathcal{E} \in \mathcal{P}(\mathcal{A})$, ($\mathcal{E} \in \texttt{Ext}(\mathcal{F}) \Leftrightarrow \mathcal{E}$ is maximal wrt $\succeq$) iff:*

1. *$\forall \mathcal{E} \in \mathcal{P}(\mathcal{A})$, if $\mathcal{E} \notin \mathcal{CF}(\mathcal{F})$ then $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ s.t. $\neg(\mathcal{E} \succeq \mathcal{E}')$*
2. *if $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$ and $\forall a' \notin \mathcal{E}$, $\exists a \in \mathcal{E}$ s.t. $a\mathcal{R}a'$, then $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ it holds that $\mathcal{E} \succeq \mathcal{E}'$*
3. *if $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$ and $\exists a' \in \mathcal{A} \setminus \mathcal{E}$ s.t. $\nexists a \in \mathcal{E}$ and $a\mathcal{R}a'$, then $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ s.t. $\neg(\mathcal{E} \succeq \mathcal{E}')$.*

In other words, a relation $\succeq$ encodes stable semantics if and only if it verifies the three conditions given in this theorem.

It is worth mentioning that there are several relations $\succeq$ that encode stable semantics. All these relations return the same maximal elements (i.e. the same extensions). However, they compare in different ways the remaining sets of arguments. An example of a relation that encodes stable semantics is the following:

Relation 1. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$. $\mathcal{E} \succeq_1 \mathcal{E}'$ iff

- $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$ and $\mathcal{E}' \notin \mathcal{CF}(\mathcal{F})$, or
- $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{F})$ and $\forall a' \in \mathcal{E}' \setminus \mathcal{E}$, $\exists a \in \mathcal{E} \setminus \mathcal{E}'$ s.t. $a\mathcal{R}a'$.

Let us illustrate this relation on the following simple example.

Example 2 *Consider the AF $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A} = \{a, b\}$ and $\mathcal{R} = \{(a, b), (b, a)\}$. It is clear that: $\{a\}$, $\{b\} \succeq_1 \{\} \succeq_1 \{a, b\}$. The two sets $\{a\}$ and $\{b\}$ are equally preferred. The maximal elements of $\succeq_1$ (its stable extensions) are $\{a\}$ and $\{b\}$.*

Note that Dung's approach returns only two classes of subsets of arguments: the extensions and the non-extensions. In Example 2, the two sets $\{a\}$ and $\{b\}$ are stable extensions while it does not say anything about the sets $\{a, b\}$ and $\{\}$. Our approach compares even the non-extensions. Indeed, according to $\succeq_1$, the set $\{\}$ is preferred to $\{a, b\}$. The fact of comparing non-extensions makes it possible to have more than one relation for stable semantics.

5. Pref-stable semantics

This section defines a new semantics, called *pref-stable*, that extends the stable one by preferences. Recall that there are two basic requirements behind stable semantics: i) conflict-freeness, and ii) external attack. The first property ensures that the extensions of a framework are conflict-free, while the second ensures that any argument outside an extension is attacked by an argument of the extension. These requirements are considered in the definition of the extensions themselves. In our approach, the requirements of pref-stable semantics are given as *postulates* that a dominance relation $\succeq$ should satisfy.

Like stable semantics, the new semantics requires that the extensions of a PAF are conflict-free wrt the attack relation. This is important since an extension represents a coherent point of view. In our approach, since all subsets of arguments are compared, we assume that a conflict-free set of arguments is preferred to any conflicting one.

Postulate 1 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ and $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{A}$. Then,*

$$\frac{\mathcal{E} \in \mathcal{CF}(\mathcal{T}) \qquad \mathcal{E}' \notin \mathcal{CF}(\mathcal{T})}{\mathcal{E} \succ \mathcal{E}'}$$ [6]

It is easy to show that if a relation satisfies this postulate, then its maximal elements are conflict-free.

Property 1 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ satisfies Postulate 1. For all $\mathcal{E} \in \mathtt{Ext}_{\succeq}(\mathcal{T})$, it holds that $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$.*

The following requirement ensures that a dominance relation is entirely based on the distinct elements of any two subsets of arguments.

Postulate 2 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF, and $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$. Then,*

$$\frac{\mathcal{E} \succeq \mathcal{E}'}{\mathcal{E} \backslash \mathcal{E}' \succeq \mathcal{E}' \backslash \mathcal{E}} \qquad \frac{\mathcal{E} \backslash \mathcal{E}' \succeq \mathcal{E}' \backslash \mathcal{E}}{\mathcal{E} \succeq \mathcal{E}'}$$

The two following postulates show how preferences between arguments are taken into account in a semantics that generalize stable semantics. As already explained, the basic idea is that if an argument a attacks another argument b and $b > a$, then the set

[6] The notation $\frac{X \quad Y}{Z}$ means that if X and Y hold, then Z holds as well.

$\{b\}$ is privileged. Thus, $\{b\}$ should be strictly preferred to $\{a\}$. However, if the two arguments are equally preferred or incomparable or even $a > b$, then the set $\{a\}$ should be strictly preferred to $\{b\}$.

The next postulate describes when a set should not be preferred to another. The idea is that: if an argument of a set $\mathcal{E}$ cannot be compared with arguments in another set $\mathcal{E}'$ (since it is neither attacked nor less preferred to any argument of the other set), then the set $\mathcal{E}$ cannot be less preferred to $\mathcal{E}'$.

Postulate 3 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF, and $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ s.t. $\mathcal{E} \cap \mathcal{E}' = \emptyset$. Then,*

$$\frac{(\exists x' \in \mathcal{E}')(\forall x \in \mathcal{E})\ \neg(x\mathcal{R}x' \wedge \neg(\ x' > x)) \wedge \neg(x > x')}{\neg(\mathcal{E} \succeq \mathcal{E}')}$$

The last postulate describes when a set is preferred to another when preferences between arguments are taken into account. The idea is that if for any argument of a set, there is at least one argument in another set which 'wins the conflict' with it, then the latter should be preferred to the former. There are two situations in which an argument x wins a conflict against x': either x attacks x' and x' does not defend itself since it is not stronger than x wrt $\geq$, or x' attacks x but x is strictly preferred to x'.

Postulate 4 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ s.t. $\mathcal{E} \cap \mathcal{E}' = \emptyset$. Then,*

$$\frac{(\forall x' \in \mathcal{E}')(\exists x \in \mathcal{E})\ s.t.\ (x\mathcal{R}x' \wedge \neg(x' > x))\ or\ (x'\mathcal{R}x \wedge x > x')}{\mathcal{E} \succeq \mathcal{E}'}$$

Now that the four postulates are introduced, we are ready to define the pref-stable semantics.

Definition 5 (Pref-stable semantics) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. A relation $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ encodes pref-stable semantics iff $\succeq$ satisfies Postulates 1, 2, 3 and 4.*

Throughout the paper, a relation that encodes pref-stable semantics will be called *pref-stable relation*, and its maximal elements are called *pref-stable extensions*.

It can be checked that a pref-stable relation strictly prefers a conflict-free set to all its strict subsets.

Property 2 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a given PAF, $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$. If $\succeq$ is a pref-stable relation, then $\mathcal{E}' \succ \mathcal{E}$ whenever $\mathcal{E} \subsetneq \mathcal{E}'$.*

Like stable semantics, there are several relations that encode pref-stable semantics. However, the differences between them are not significant, and we can show that they all return the same pref-stable extensions.

Theorem 2 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF. If $\succeq, \succeq' \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ are pref-stable relations, then $\mathtt{Ext}_{\succeq}(\mathcal{T}) = \mathtt{Ext}_{\succeq'}(\mathcal{T})$.*

Finally, we can show that a pref-stable semantics generalizes stable semantics. This means that when preferences are not available or do not conflict with attacks in a given PAF, then pref-stable relations are a subset of those encoding stable semantics (i.e. they satisfy the three conditions of Theorem 1.

Theorem 3 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ its basic version. If $\succeq$ is a pref-stable relation and $\nexists a, b \in \mathcal{A}$ such that $a\mathcal{R}b$ and $b > a$, then:*

- $\mathtt{Ext}(\mathcal{F}) = \mathtt{Ext}_{\succeq}(\mathcal{T})$
- *$\succeq$ satisfies the three conditions of Theorem 1*

Let us now consider an example of a pref-stable relation. This relation extends $\succeq_1$ which encodes stable semantics.

Relation 2 (Relation 1 extended). Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$. $\mathcal{E} \succeq_2 \mathcal{E}'$ iff at least one of the following conditions holds:

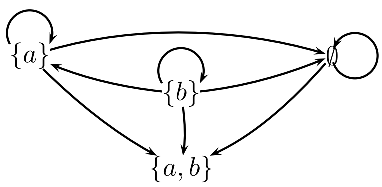
- $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ and $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$
- $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ and $(\forall a' \in \mathcal{E}' \setminus \mathcal{E})(\exists a \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $(a\mathcal{R}a' \wedge a' \not> a) \vee (a > a')$.

Property 3 *$\succeq_2$ is a pref-stable relation.*

Let us illustrate this relation on the following simple example.

Example 3 *Let $\mathcal{A} = \{a, b\}$, $\mathcal{R} = \{(a, b)\}$ and $b \geq a$. It can be checked that the set $\{b\}$ is the only maximal element of relation $\succeq_2$. Figure 1 shows the preferences among elements of $\mathcal{P}(\mathcal{A})$ wrt $\succeq_2$.*

Figure 1. $\succeq_2 \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$



Let us now reconsider the example presented in the introduction.

Example 1 (Cont): It can be checked that every pref-stable relation returns exactly two pref-extensions: $\{a_1, a_2, a_4\}$ (whose base is $\{x, \neg y\}$) and $\{a_1, a_3, a_6\}$ (whose base is $\{x, x \rightarrow y\}$). Thus, the bases corresponding to both extensions are consistent.

5.1. General and specific pref-stable relations

As already said, there are several relations that encode pref-stable semantics. The aim of this section is to define the upper and lower bounds of these relations.

The following relation, denoted by $\succeq_g$, is the most general pref-stable relation. It returns $\mathcal{E} \succeq_g \mathcal{E}'$ if and only if it can be proved from the four postulates that $\mathcal{E}$ must be preferred to $\mathcal{E}'$.

Definition 6 (General pref-stable relation) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$. $\mathcal{E} \succeq_g \mathcal{E}'$ iff:*

- *$\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ and $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$, or*
- *$\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ and $(\forall a' \in \mathcal{E}' \setminus \mathcal{E})(\exists a \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $(a\mathcal{R}a' \wedge a' \not> a) \vee (a'\mathcal{R}a \wedge a > a')$.*

Property 4 *$\succeq_g$ is a pref-stable relation.*

The next relation, denoted by $\succeq_s$, is the most specific pref-stable relation. It returns $\mathcal{E} \succeq_s \mathcal{E}'$ if and only if from the four postulates it cannot be proved that $\neg(\mathcal{E} \succeq_s \mathcal{E}')$.

Definition 7 (Specific pref-stable relation) *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$. $\mathcal{E} \succeq_s \mathcal{E}'$ iff:*

- *$\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$, or*
- *$\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ and $(\forall a' \in \mathcal{E}' \setminus \mathcal{E})(\exists a \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $(a\mathcal{R}a' \wedge a' \not> a) \vee (a > a')$.*

Property 5 *$\succeq_s$ is a pref-stable relation.*

Let us illustrate the differences between the three relations $\succeq_2$, $\succeq_s$ and $\succeq_g$ on the following example.

Example 4 *Let $\mathcal{A} = \{a, b, c\}, \mathcal{R} = \{(a, b)\}$ and $a \geq c$. For example, it holds that $\{a\} \succeq_2 \{c\}$, $\{a\} \succeq_s \{c\}$ and $\neg(\{a\} \succeq_g \{c\})$. That is, for relations $\succeq_2$ and $\succeq_s$ the strict preference between a and c is enough to prefer $\{a\}$ to $\{c\}$. For relation $\succeq_g$, since c is not attacked by a, there is no preference between the sets $\{a\}$ and $\{c\}$. The fact that a is stronger is not important, because there is no conflict between those arguments.*

Another difference is that for the relation $\succeq_s$, all conflicting sets are equally preferred. For example, $\{a, b, c\} \succeq_s \{a, b\}$ and $\{a, b\} \succeq_s \{a, b, c\}$. Besides, relations $\succeq_2$ and $\succeq_g$ encode the idea that a contradictory point of view cannot be accepted as a standpoint. Thus, it is not even possible to compare two contradictory sets of arguments. For example $\neg(\{a, b, c\} \succeq_2 \{a, b\})$.

The next result shows that any pref-stable relation is "between" the general and the specific relations.

Theorem 4 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$. Let $\succeq$ be a pref-stable relation.*

- *If $\mathcal{E} \succeq_g \mathcal{E}'$ then $\mathcal{E} \succeq \mathcal{E}'$.*
- *If $\mathcal{E} \succeq \mathcal{E}'$ then $\mathcal{E} \succeq_s \mathcal{E}'$.*

A simple consequence of the previous result is that, if $\mathcal{E} \succeq_g \mathcal{E}'$ and $\mathcal{E} \succeq_s \mathcal{E}'$, then $\mathcal{E} \succeq \mathcal{E}'$ for any pref-stable relation.

5.2. Corresponding Semantics

This section characterizes pref-stable extensions without referring to pref-stable relations. Indeed, the next theorem proves that it is not necessary to compare all sets of arguments in order to know whether a given subset of arguments is a pref-stable extension of a PAF.

Theorem 5 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\succeq$ a pref-stable relation. $\mathcal{E} \in \mathtt{Ext}_{\succeq}(\mathcal{T})$ iff:*

- *$\mathcal{E} \in \mathcal{CF}(\mathcal{T})$, and*
- *$(\forall a' \in \mathcal{A} \setminus \mathcal{E})$ $(\exists a \in \mathcal{E})$ s.t. $(a\mathcal{R}a' \wedge a' \not> a) \vee (a'\mathcal{R}a \wedge a > a')$.*

This result is of great importance since it shows how to compute directly the pref-stable extensions of a PAF without bothering about pref-stable relations. This is particularly the case when we do not want to compare all the elements of $\mathcal{P}(\mathcal{A})$.

Another way to compute the pref-stable extensions of a PAF is to "invert" the direction of attacks when they are not in accordance with the preferences between arguments. We apply then stable semantics on the basic framework that is obtained. More precisely, we start with a PAF $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$. We compute an AF $\mathcal{F} = (\mathcal{A}, \mathcal{R}')$ where $\mathcal{R}'$ is defined as follows:

$$\begin{cases} \text{If } (a,b) \in \mathcal{R} \text{ and } b \not> a \text{ then } (a,b) \in \mathcal{R}' \\ \text{If } (a,b) \in \mathcal{R} \text{ and } b > a \text{ then } (b,a) \in \mathcal{R}' \end{cases}$$

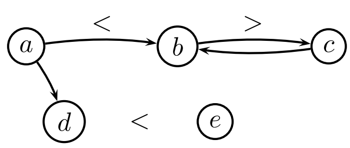
then we apply stable semantics on the new framework $(\mathcal{A}, \mathcal{R}')$. This result is proved in the following theorem.

Theorem 6 *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF and $\succeq$ be a pref-stable relation. Let $\mathcal{R}' = \{(a,b) \mid (a\mathcal{R}b \wedge b \not> a) \vee (b\mathcal{R}a \wedge a > b)\}$. It holds that $\mathtt{Ext}_{\succeq}(\mathcal{T}) = \mathtt{Ext}((\mathcal{A}, \mathcal{R}'))$.*

Let us illustrate this result through a simple example.

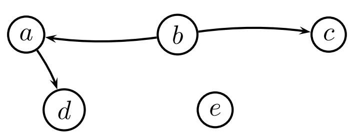
Example 5 *Let us consider the PAF represented in Figure 2. It can be checked that any pref-stable relation will return exactly one pref-stable extension: $\mathtt{Ext}_{\succeq}(\mathcal{T}) = \{\{b, d, e\}\}$. The argumentation framework that is obtained after inverting arrows is de-*

Figure 2. PAF $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ (Example 5)



picted in Figure 3. It is easy to see that the only stable extension of this framework is the set $\{b, d, e\}$.

Figure 3. Framework $(\mathcal{A}, \mathcal{R}')$



6. Conclusion

Several proposals have been made in the literature on how to integrate preferences in an argumentation system. In this paper, we have shown that those proposals may return undesirable results when the attack relation is asymmetric. We have then proposed a novel approach to compute the extensions of a PAF. The idea is to define new acceptability semantics that take into account both attacks and preferences between arguments.

In our approach, a semantics is defined by a dominance relation on the powerset of the set of arguments. The extensions of a PAF are the maximal elements of this relation. This approach offers great advantages. First, it shows clearly the impact of preferences on the result of a PAF. Second, it allows to compare all the elements of the powerset of arguments. Thus, it offers more information.

In this paper, we have mainly focused on generalizing Dung's semantics [8], in particular stable one. We have defined a new semantics, called pref-stable, that recovers stable semantics in case preferences between arguments are not available or do not conflict with the attacks. We have proposed a full characterization of pref-stable semantics both in terms of dominance relations that encode it and also in a declarative way.

To the best of our knowledge, the only related work is that proposed in [3]. In that paper, three "particular" relations that extend respectively stable, preferred and grounded semantics are provided. As shown in our paper, those relations are unfortunately not unique. We have provided a full picture on the way of extending stable semantics into pref-stable using postulates.

References

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

[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 *Proceedings of IJCAI'09*, pages 665–670. AAAI Press, 2009.

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

[5] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of UAI'93*, pages 411–419, 1993.

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

[7] C. Cayrol, V. Royer, and C. Saurel. Management of preferences in assumption-based reasoning. *Lecture Notes in Computer Science*, 682:13–22, 1993.

[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] M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as 'logical uncertainty'. In *Proceedings of ECSQARU'93*, number 747, pages 85–90. 1993.

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

[11] G. Simari and R. Loui. A mathematical treatment of defeasible reasoning and its implementation. *AIJ*, 53:125–157, 1992.

Appendix

Proof of Property 1 Let us assume that $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. Thus, $\mathcal{E} \succeq \emptyset$. Since $\emptyset \in \mathcal{CF}(\mathcal{T})$, then from Postulate 1, $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$.

Proof of Property 2 From Postulate 2, it follows that $\mathcal{E} \succeq \mathcal{E}'$ iff $\emptyset \succeq \mathcal{E}' \setminus \mathcal{E}$. From Postulate 3, $\neg(\emptyset \succeq \mathcal{E}' \setminus \mathcal{E})$. Consequently $\neg(\mathcal{E} \succeq \mathcal{E}')$. Postulate 4 implies that $\mathcal{E}' \setminus \mathcal{E} \succeq \emptyset$. From Postulate 2, $\mathcal{E}' \succeq \mathcal{E}$. This fact, together with $\neg(\mathcal{E} \succeq \mathcal{E}')$, leads to conclusion $\mathcal{E}' \succ \mathcal{E}$.

Proof of Property 3 To show that $\succeq_2$ is a pref-stable relation, we show that it satisfies all postulates. Postulate 1 is satisfied since from the first item of the definition of $\succeq_2$, any conflict-free set is preferred to any conflicting set. Postulate 2 is satisfied since from the second item of the same definition, when comparing two sets $\mathcal{E}$ and $\mathcal{E}'$, common elements are not taken into account. The second condition of the definition of $\succeq_2$ is exactly the negation of the condition of Postulate 3. Since Postulate 4 implies the second item of this definition, then it is verified.

Proof of Property 4 Postulates 1 and 2 are verified for the same reasons as for $\succeq_2$. Postulate 3 implies that the second item of Definition 6 is not satisfied. Postulate 4 is trivially verified.

Proof of Property 5 We see from the first item of Definition 7 that all (conflict-free and non conflict-free) sets are better than non conflict-free sets. A non conflict-free set, however, cannot be better than conflict-free set. Thus, Postulate 1 is satisfied. Postulates 2, 3 and 4 are verified for same reasons as in the case of relation $\succeq_g$.

Proof of Theorem 1. $\Rightarrow$ Assume that $((\forall \mathcal{E} \subseteq \mathcal{A})$ we have $\mathcal{E} \in \texttt{Ext}(\mathcal{F}) \Leftrightarrow \mathcal{E}$ is a maximal element of $\succeq)$. We will prove that the three above conditions are satisfied.

1. Assume that $\mathcal{E} \subseteq \mathcal{A}$ and $\mathcal{E} \notin \mathcal{CF}(\mathcal{F})$. So, $\mathcal{E}$ is not a stable extension of $(\mathcal{A}, \mathcal{R})$. From what we supposed, $\mathcal{E}$ is not a maximal element of $\succeq$. In other words, $\exists \mathcal{E}' \subseteq \mathcal{A}$ s.t. $\neg(\mathcal{E} \succeq \mathcal{E}')$.
2. Assume that $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$ and that $\forall a' \notin \mathcal{E}$, $\exists a \in \mathcal{E}$ s.t. $(a, a') \in \mathcal{R}$. Thus, $\mathcal{E}$ is a Dung's stable extension of $(\mathcal{A}, \mathcal{R})$. From what we supposed, it must be that $\mathcal{E}$ is a maximal element of $\succeq$. Consequently, $(\forall \mathcal{E}' \subseteq \mathcal{A})\mathcal{E} \succeq \mathcal{E}'$.
3. Assume that $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$ and $\exists a' \in \mathcal{A} \setminus \mathcal{E}$ s.t. $\nexists a \in \mathcal{E}$ and $(a, a') \in \mathcal{R}$. It is obvious that $\mathcal{E}$ is not a Dung's stable extension of $(\mathcal{A}, \mathcal{R})$. From $(\mathcal{E} \in \texttt{Ext}(\mathcal{F}) \Leftrightarrow \mathcal{E}$ is a maximal element of $\succeq)$ we conclude that $\mathcal{E}$ is not a maximal element of $\succeq$. Thus, $(\exists \mathcal{E}' \subseteq \mathcal{A})\neg(\mathcal{E} \succeq \mathcal{E}')$.

$\Leftarrow$ Let $\succeq$ satisfy the three conditions.

- Let $\mathcal{E}$ be a stable extension of $(\mathcal{A}, \mathcal{R})$ and let $\mathcal{E}' \subseteq \mathcal{A}$. From the second condition, $\mathcal{E} \succeq \mathcal{E}'$. Thus, $\mathcal{E}$ must be a maximal element wrt $\succeq$.
- If $\mathcal{E}$ is not a stable extension of $(\mathcal{A}, \mathcal{R})$ but $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$, from the third condition we have that $\mathcal{E}$ is not a maximal element wrt $\succeq$.
- If $\mathcal{E}$ is not a stable extension of $(\mathcal{A}, \mathcal{R})$ and $\mathcal{E} \notin \mathcal{CF}(\mathcal{F})$ then, from the first condition, $\mathcal{E}$ is not a maximal element wrt $\succeq$.

Proof of Theorem 2 $\Rightarrow$ Let $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. We will prove that $\mathcal{E} \in \texttt{Ext}_{\succeq'}(\mathcal{T})$. From Postulate 1, $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$. Let $\mathcal{E}' \subseteq \mathcal{A}$. If $\mathcal{E}'$ is not conflict-free then, from Postulate 1, $\mathcal{E} \succeq' \mathcal{E}'$. Else, from Postulate 2, $\mathcal{E} \succeq' \mathcal{E}'$ iff $\mathcal{E} \setminus \mathcal{E}' \succeq' \mathcal{E}' \setminus \mathcal{E}$. Let $\mathcal{E}_1 = \mathcal{E} \setminus \mathcal{E}'$ and $\mathcal{E}_2 = \mathcal{E}' \setminus \mathcal{E}$. $\mathcal{E}_1$ and $\mathcal{E}_2$ are disjunct conflict-free sets. If condition of Postulate 4 is satisfied for $\mathcal{E}_1$ and $\mathcal{E}_2$, then $\mathcal{E}_1 \succeq' \mathcal{E}_2$. Let us study the case when this condition is not satisfied. Condition of Postulate 3 is not satisfied since $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. Thus, it must be that $(\exists x' \in \mathcal{E}_2)$ s.t. $(\nexists x \in \mathcal{E}_1)((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >)$ and $(\exists x \in \mathcal{E}_1)(x, x') \in >$. Let $X = \{x \in \mathcal{E}_1 | (x, x') \in >\}$. X is conflict-free. From Postulate 3, $\neg(\mathcal{E}_1 \setminus X \succeq \{x'\})$. Postulate 2 implies that $\neg(\mathcal{E}_1 \setminus X \cup (X \cup (\mathcal{E} \cap \mathcal{E}')) \succeq \{x'\} \cup (X \cup (\mathcal{E} \cap \mathcal{E}')))$, i.e. $\neg(\mathcal{E} \succeq \{x'\} \cup (X \cup (\mathcal{E} \cap \mathcal{E}')))$. Contradiction with $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. Thus, condition of Postulate 4 is satisfied for $\mathcal{E}_1$ and $\mathcal{E}_2$, and $\mathcal{E}_1 \succeq' \mathcal{E}_2$. Consequently, $\mathcal{E} \succeq' \mathcal{E}'$. This means that $\mathcal{E} \in \texttt{Ext}_{\succeq'}(\mathcal{T})$.
$\Leftarrow$ In the first part of proof, we showed that for all pref-stable relations $\succeq_1, \succeq_2$, it holds that if $\mathcal{E} \in \texttt{Ext}_{\succeq_1}(\mathcal{T})$ then $\mathcal{E} \in \texttt{Ext}_{\succeq_2}(\mathcal{T})$. Contraposition of this rule gives if $\mathcal{E} \notin \texttt{Ext}_{\succeq_2}(\mathcal{T})$ then $\mathcal{E} \notin \texttt{Ext}_{\succeq_1}(\mathcal{T})$. Since this was proved for arbitrary relations which satisfy all postulates, we conclude: if $\mathcal{E} \notin \texttt{Ext}_{\succeq}(\mathcal{T})$ then $\mathcal{E} \notin \texttt{Ext}_{\succeq'}(\mathcal{T})$.

Proof of Theorem 3

- Let $\mathcal{T}$ be a preference-based argumentation system s.t. $(\nexists a, b \in \mathcal{A})(a, b) \in \mathcal{R} \wedge (b, a) \in >$.
$\Rightarrow$ Let $\mathcal{E} \in \texttt{Ext}(\mathcal{F})$. We prove that $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. Let $\mathcal{E}' \in \mathcal{P}(\mathcal{A})$. If $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$ then, from Postulate 1, $\mathcal{E} \succeq \mathcal{E}'$. Let $\mathcal{E}' \in \mathcal{CF}(\mathcal{T})$. Since $\mathcal{E} \in \texttt{Ext}(\mathcal{F})$ then $(\forall x' \in \mathcal{E}' \backslash \mathcal{E})(\exists x \in \mathcal{E} \backslash \mathcal{E}')(x, x') \in \mathcal{R}$. We supposed $(\nexists a, b \in \mathcal{A})(a, b) \in \mathcal{R} \wedge (b, a) \in >$. Thus, from Postulate 4, $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$. Now, Postulate 2 implies $\mathcal{E} \succeq \mathcal{E}'$. Since $\mathcal{E}'$ was arbitrary, then $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$.
$\Leftarrow$ Let $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. We will show that $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{F})$. From Postulate 1, $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$. Let $x' \notin \mathcal{E}$. Since $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$ then it must be $\mathcal{E} \succeq \{x'\}$. From Postulate 3, $(\exists x \in \mathcal{E})(x, x') \in \mathcal{R} \vee (x, x') \in >$. If $(\exists x \in \mathcal{E})(x, x') \in \mathcal{R}$, the proof is over. Let us suppose the contrary. Then $(\nexists x \in \mathcal{E})(x, x') \in \mathcal{R}$. Let $X = \{x \in \mathcal{E} | x > x'\}$. From Postulate 3, $\neg(\mathcal{E} \setminus X \succeq \{x'\})$. This fact and Postulate 2 imply $\neg(\mathcal{E} \succeq (X \cup \{x'\}))$. Contradiction with $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$. Thus, $\mathcal{E} \in \texttt{Ext}(\mathcal{F})$.
- In the first part of the proof, we have shown that for every PAF $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ s.t. $(\nexists a, b \in \mathcal{A}) a\mathcal{R}b \wedge b > a$ and for every pref-relation $\succeq$, it holds that maximal elements wrt $\succeq$ are exactly stable extensions of argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$. Informally speaking, $\succeq$ generalizes stable semantics. Formally, from the fact that for any $\mathcal{E} \subseteq \mathcal{A}$ is holds that ($\mathcal{E}$ is a maximal element wrt $\succeq$ iff $\mathcal{E} \in \texttt{Ext}(\mathcal{F})$), Theorem 1 implies that three items of that theorem must be verified by $\succeq$.

Proof of Theorem 4

- Let $\mathcal{E} \succeq_g \mathcal{E}'$. This means that $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$. If $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$, then from Postulate 1, $\mathcal{E} \succeq \mathcal{E}'$. We study the case when $\mathcal{E}' \in \mathcal{CF}(\mathcal{T})$. From Postulate 2, we have $\mathcal{E} \succeq \mathcal{E}'$ iff $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$. From Definition 6 and Postulate 4, $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$. Thus, $\mathcal{E} \succeq \mathcal{E}'$.
- If $\mathcal{E}, \mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$ then, Definition 7 implies $\mathcal{E} \succeq_s \mathcal{E}'$. Case $\mathcal{E} \notin \mathcal{CF}(\mathcal{T})$, $\mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ is not possible because of Postulate 1. If $\mathcal{E} \in \mathcal{CF}(\mathcal{T}), \mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$, then from Definition 7, $\mathcal{E} \succeq_s \mathcal{E}'$. In the non-trivial case, when $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$, from Postulate 2, $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$. Suppose that $\neg(\mathcal{E} \setminus \mathcal{E}' \succeq_s \mathcal{E}' \setminus \mathcal{E})$. Now, Definition 7 implies $(\exists x' \in \mathcal{E}' \setminus \mathcal{E})(\nexists x \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $((x, x') \in >) \vee ((x, x') \in \mathcal{R} \wedge (x', x) \notin >)$. From this fact and Postulate 3, it holds that $\neg(\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E})$. Contradiction.

Proof of Theorem 5 Since both relations $\succeq$ and $\succeq_g$ verify Postulates 1, 2, 3 and 4, then from Theorem 2, $\texttt{Ext}_{\succeq}(\mathcal{T}) = \texttt{Ext}_{\succeq_g}(\mathcal{T})$. This means that it is sufficient to prove that $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$ iff the two conditions of theorem are satisfied.
$\Rightarrow$ Let $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$. Since $\mathcal{E}$ is a pref-extension, according to Property 1, $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$. Let $x' \in \mathcal{A} \setminus \mathcal{E}$. We supposed that $(\nexists a \in \mathcal{A})$ s.t. $(a, a) \in \mathcal{R}$, so it must be that $\{x'\}$ is conflict-free. Since $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$, it holds that $\mathcal{E} \succeq_g \{x'\}$. Since $\mathcal{E}$ and $\{x'\}$ are conflict-free, Definition 6 implies $(\exists x \in \mathcal{E})$ s.t. $(((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >))$.
$\Leftarrow$ Let $\mathcal{E}$ be conflict-free set and let $(\forall x' \in \mathcal{A} \backslash \mathcal{E})$ $(\exists x \in \mathcal{E})$ s.t. $(((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >))$. Let us prove that $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$.

- Since $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ then for every non conflict-free set $\mathcal{E}'$ it holds that $\mathcal{E} \succeq_g \mathcal{E}'$.
- Let $\mathcal{E}' \subseteq \mathcal{A}$ be an arbitrary conflict-free set of arguments. If $\mathcal{E}' \subseteq \mathcal{E}$, the second condition of theorem is trivially satisfied. Else, let $x' \in \mathcal{E}' \setminus \mathcal{E}$. From what we supposed, we have that $(\exists x \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $((x, x') \in \mathcal{R} \wedge (x', x) \notin >)$ or $((x', x) \in \mathcal{R} \wedge (x, x') \in >)$. Thus, $\mathcal{E} \succeq_g \mathcal{E}'$.

From those two items, we have that $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$.

Proof of Theorem 6 Since both relations $\succeq$ and $\succeq_g$ verify Postulates 1, 2, 3 and 4, then from Theorem 2, $\texttt{Ext}_{\succeq}(\mathcal{T}) = \texttt{Ext}_{\succeq_g}(\mathcal{T})$. This means that it is sufficient to prove that $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$ iff $\mathcal{E} \in \texttt{Ext}((\mathcal{A}, \mathcal{R}'))$. Note also that $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ iff $\mathcal{E}$ is conflict-free in $(\mathcal{A}, \mathcal{R}')$. Thus, we will simply use the notation $\mathcal{E} \in \mathcal{CF}$ to refer to both of those cases since they coincide.
$\Rightarrow$ Let $\mathcal{E} \in \texttt{Ext}_{\succeq_g}(\mathcal{T})$. From Theorem 5, $\mathcal{E} \in \mathcal{CF}$ and $(\forall x' \in \mathcal{A} \setminus \mathcal{E})$ $(\exists x \in \mathcal{E})$ s.t. $(((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >))$. This means that $(\forall x' \in \mathcal{A} \setminus \mathcal{E})$ $(\exists x \in \mathcal{E})$ s.t. $(x, x') \in \mathcal{R}'$. In other words, $\mathcal{E} \in \texttt{Ext}((\mathcal{A}, \mathcal{R}'))$.
$\Leftarrow$ Let $\mathcal{E} \in \texttt{Ext}((\mathcal{A}, \mathcal{R}'))$. Trivially, $\mathcal{E} \in \mathcal{CF}$. Let $\mathcal{E}' \subseteq \mathcal{A}$. If $\mathcal{E}' \notin \mathcal{CF}$, then $\mathcal{E} \succeq \mathcal{E}'$. Else, let $\mathcal{E}' \in \mathcal{CF}$. Since $\mathcal{E} \in \texttt{Ext}((\mathcal{A}, \mathcal{R}'))$, then $(\forall x' \in \mathcal{A} \setminus \mathcal{E})(\exists x \in \mathcal{E})(x, x') \in \mathcal{R}'$. This is equivalent to $(\forall x' \in \mathcal{A} \setminus \mathcal{E})$ $(\exists x \in \mathcal{E})$ s.t. $(((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >))$. Trivially, $(\forall x' \in \mathcal{E}' \setminus \mathcal{E})$ $(\exists x \in \mathcal{E} \setminus \mathcal{E}')$ s.t. $(((x, x') \in \mathcal{R} \wedge (x', x) \notin >) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in >))$. That means that $\mathcal{E} \in \texttt{Ext}_{\succeq}(\mathcal{T})$.