

# A new approach for preference-based argumentation frameworks

Leila Amgoud · Srdjan Vesic

the date of receipt and acceptance should be inserted later

**Abstract** Dung's argumentation framework consists of a set of *arguments* and an *attack* relation among them. Arguments are evaluated and *acceptable* sets of them, called *extensions*, are computed using a given *semantics*. Each extension represents a coherent position. In the literature, several proposals have extended this framework in order to take into account the strength of arguments. The basic idea is to ignore an attack if the attacked argument is stronger than (or preferred to) its attacker. Semantics are then applied using only the remaining attacks.

In this paper, we show that those proposals behave correctly when the attack relation is symmetric. However, when it is asymmetric, *conflicting extensions* may be computed leading to unintended conclusions. We propose an approach that guarantees conflict-free extensions. This approach presents two novelties: the first one is that it takes into account preferences at the semantics level rather than the attack level. The idea is to extend existing semantics with preferences. 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. The second novelty of our approach is that a semantics is defined as a *dominance relation* on the powerset of the set of arguments. The extensions (under a semantics) are the maximal elements of the dominance relation. Such an approach makes it possible not only to compute the extensions of a framework but also to compare its non-extensions. We start by proposing three dominance relations that generalize respectively stable, preferred and grounded semantics with preferences. Then, we focus on stable semantics and provide full characterizations of its dominance relations and those of its generalized versions. Complexity results are provided. Finally, we show that an instance of the proposed framework retrieves the preferred sub-theories which were proposed in the context of handling inconsistency in weighted knowledge bases.

**Keywords** Argumentation · Preferences

---

Leila Amgoud · Srdjan Vesic  
IRIT - CNRS, 118 route de Narbonne, 31062 Toulouse Cedex 9, France  
E-mail: amgoud@irit.fr

## 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, or to choose an option, etc. Due to its explanatory power, argumentation has gained increasing interest in Artificial Intelligence, namely for handling inconsistency (e.g. [5, 19, 46]), making decisions (e.g. [10, 20]), learning concepts (e.g. [11, 40]) and modeling different types of dialogues (e.g. [9, 36, 42, 47]).

Most of the models that treat the cited applications are instantiations of an abstract framework developed in [29]. This framework consists of a set of *arguments* and a binary relation that captures *attacks* among them. Different acceptability *semantics* were proposed in the same paper. A semantics defines sets of acceptable arguments, called *extensions*. Each extension is intended to represent a “coherent” position (or point of view). This notion of coherence is captured by *conflict-freeness* in acceptability semantics. Indeed, a basic requirement for each semantics is that it ensures conflict-free extensions, that is each extension does not contain an argument and its attacker. Note that violating conflict-freeness leads to counter-intuitive results meaning that the framework fails to meet its objectives. Let us assume that an argumentation framework is built for reasoning with defeasible/inconsistent information gathered in a knowledge base. In this case, attacks capture the inconsistency of the knowledge. If conflict-freeness is violated, then the framework may have extensions that support contradictory conclusions, and thus it may accept a conclusion and its contrary. This means that the framework collapses.

There is a clear consensus in the literature that arguments do not necessarily have the same strength. It may be the case that an argument relies on certain information while another argument is built from less certain ones, or that an argument promotes an important value while another promotes a weaker one. In both cases, the former argument is clearly stronger than the latter. These differences in arguments’ strengths make it possible to compare them. Consequently, several preference relations between arguments have been defined in the literature (e.g. [7, 18, 27, 44, 46]). There is also a consensus on the fact that preferences should be taken into account in the evaluation of arguments (see [6, 17, 38, 44, 46]).

Surprisingly, there are divergent opinions on whether the attack relation in Dung’s framework already takes into account the strengths of arguments or should be augmented by a preference relation which captures these strengths. It is worth mentioning that in [29], there is no answer to this question. The only thing which is mentioned in that paper is that an argument can attack another argument meaning that it disqualifies this argument, and the two arguments cannot survive together.

According to some researchers like Kaci [33, 35], the attack relation in Dung’s framework is a combination of a *symmetric* conflict relation and a preference relation between arguments. The author argues that a conflict between two arguments should always be symmetric, and since Dung’s attack relation may be asymmetric, this means that a preference relation is applied between the two arguments in order to solve the conflict.

According to other researchers [4], an argument can attack another argument by undermining one of its three basic components, that is its conclusion, a premise of its support, or a link between a premise and a conclusion. The *formal* definition of the first kind of attack induces a symmetric relation (the so-called *rebut* [32]), the two other kinds of attack induce asymmetric relations called respectively *assumption*

*attack* [32] and *undercut* [41]. Thus, the conflict relation mentioned by Kaci can be either symmetric or asymmetric. Besides, in a recent paper [1], Amgoud and Besnard have shown that the choice of an attack relation is crucial for ensuring sound results, and should not be arbitrary. They have studied how to choose an attack relation when arguments are built using any logic satisfying Tarski's axioms [48]. The results confirm that an attack relation *should not* be symmetric, in particular when the knowledge base from which arguments are built contains at least one minimal inconsistent subset with a cardinality higher than 2. Indeed, symmetric relations lead to the violation of the rationality postulates identified in [24]. This means that the point of view defended by Kaci is wrong, and thus Dung's framework should be extended by preferences at least for those applications which use Tarski's logic for building arguments.

In [6], a first *abstract* preference-based argumentation framework was proposed. It takes as input a set of arguments, an attack relation, and a preference relation between arguments which is abstract and can be instantiated in different ways. This proposal was generalized in [38] in order to reason even about preferences. Thus, arguments may support preferences about arguments. Another work which considers preferences between arguments is that proposed in [37]. This work is largely inspired from [6] and its notion of strict defeat. The last extension, proposed in [17], 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 applied on the remaining attacks.

In this paper, we show that while the above idea seems meaningful, the existing frameworks (namely [6,17,38]) may return conflicting extensions leading thus to unintended results when the attack relation is *asymmetric*. Then, we propose a new approach that guarantees conflict-freeness of extensions. This approach introduces two novelties. The first one concerns the level at which it takes into account preferences. Contrarily to existing approaches which combine the preference relation with the attack relation and produce a new relation, called *defeat*, the solution we propose in this paper takes into account preferences at the semantics level. The idea is thus to define new acceptability semantics that are based on the attack and preference relations. In case preferences are not available or do not conflict with the attacks, the extensions of the new semantics should coincide with those of the basic semantics proposed by Dung in [29]. Our aim is not to define new semantics but to improve existing ones by preferences.

The second novelty of our approach is that a semantics is defined as a *dominance relation* on the powerset of the set of arguments. The extensions (under such a semantics) are the maximal elements of the corresponding dominance relation. Thus, contrarily to existing semantics which partition the powerset of the set of arguments into two subsets: the extensions and the non-extensions, our approach provides more information since it compares all the subsets of arguments. Such information may be useful for instance in dialogues in order to compare the sets of arguments advanced by agents, or in a decision making problem for ordering a set of alternatives instead of returning only the best one (e.g. [8]).

The paper contains two main parts: in the first part we propose three dominance relations that improve respectively grounded, preferred and stable semantics with pref-

ferences. In the second part of the paper we focus only on stable semantics. We show that it can be improved by different dominance relations. We provide a full characterization of these dominance relations through a set of postulates. The most general and the most specific dominance relations are identified. We also show that the differences between those relations are not significant, as we prove that they all return the same maximal elements (which correspond to extensions). We provide theorems which describe the extensions of the new semantics without referring to any dominance relation. Then, we show that the computation complexity of different reasoning tasks remains the same as in the basic framework (i.e. without preferences). Finally, we show that an instance of our PAF recovers the preferred sub-theories proposed by Brewka in [21] for handling inconsistency in prioritized knowledge bases.

The paper is organized as follows: Section 2 recalls Dung’s argumentation framework as well as its extensions with preferences. Section 3 shows through an example the limits of existing PAFs. Section 4 introduces our new approach. Section 5 defines three new acceptability semantics that generalize respectively stable, preferred and grounded semantics with preferences. Section 6 deeply studies the different ways of generalizing stable semantics. Section 7 shows how an instance of the PAF computes the preferred sub-theories of a prioritized knowledge base. Section 8 compares our approach with existing works. Section 9 is devoted to some concluding remarks and perspectives. The last section is an appendix containing the proofs or all our results.

## 2 Dung’s framework and its extensions

In the seminal paper [29], an *argumentation framework* (AF) is a pair  $(\mathcal{A}, \mathcal{R})$ , where  $\mathcal{A}$  is a set of *arguments* and  $\mathcal{R}$  is a binary relation between arguments representing *attacks* among them ( $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ ). The notation  $(a, b) \in \mathcal{R}$  or  $a\mathcal{R}b$  means that the argument  $a$  attacks the argument  $b$ . Different acceptability *semantics* have been proposed in the same paper for evaluating arguments. Each semantics describes what are the “good” subsets of arguments, called *extensions*. Depending on the semantics, an argumentation framework may have none or one or several extensions. An extension represents a *coherent position* or a *coherent point of view*. Thus, it should be conflict-free (i.e. it does not contain two conflicting arguments) and it should defend its elements. These two requirements are formally defined below.

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

- $\mathcal{B}$  is *conflict-free* iff  $\nexists a, b \in \mathcal{B}$  such that  $a\mathcal{R}b$ .
- $\mathcal{B}$  *defends* an argument  $a$  iff  $\forall b \in \mathcal{A}$  if  $b\mathcal{R}a$ , then  $\exists c \in \mathcal{B}$  such that  $c\mathcal{R}b$ .

The main semantics introduced in [29] are recalled below.

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

- $\mathcal{B}$  is *admissible* iff it is conflict-free and defends all its elements.
- $\mathcal{B}$  is a *preferred extension* iff it is a maximal (wrt set  $\subseteq$ ) admissible set.
- $\mathcal{B}$  is a *stable extension* iff  $\mathcal{B}$  is conflict-free and  $\forall a \in \mathcal{A} \setminus \mathcal{B}, \exists b \in \mathcal{B}$  such that  $b\mathcal{R}a$ .
- $\mathcal{B}$  is a *grounded extension* iff  $\mathcal{B}$  is the least fixpoint of a function  $\mathcal{C}$  where  $\mathcal{C}(\mathcal{S}) = \{a \in \mathcal{A} \mid \mathcal{S} \text{ defends } a\}$ , for  $\mathcal{S} \subseteq \mathcal{A}$ .

Notation: Let  $\mathcal{F} = (\mathcal{A}, \mathcal{R})$  be an AF.  $\text{Ext}_x(\mathcal{F})$  denotes the set of extensions of  $\mathcal{F}$  under semantics  $x$ . When the semantics is clear from the context, we will use  $\text{Ext}(\mathcal{F})$  for short.

Several other acceptability semantics have been proposed in the literature (for instance, semi-stable semantics [25], ideal semantics [30], recursive semantics [16]). It is worth mentioning that conflict-freeness is at the heart of all those new semantics. For the purpose of this paper, we only focus on Dung’s semantics.

In [6], a first extension of Dung’s framework has been proposed. It takes as input a set  $\mathcal{A}$  of arguments, an attack relation  $\mathcal{R}$ , and a (partial or total) preorder (i.e. a *reflexive* and *transitive* binary relation)  $\geq$  on  $\mathcal{A}$ . This preorder represents a preference relation between arguments. The expression  $(a, b) \in \geq$  or  $a \geq b$  means that the argument  $a$  is at least as *strong* as (or as *preferred* as)  $b$ . The symbol  $>$  denotes the strict relation associated with  $\geq$ . Indeed,  $a > b$  iff  $a \geq b$  and not  $(b \geq a)$ . Finally, when two arguments are equally preferred (i.e.  $a \geq b$  and  $b \geq a$ ), they are denoted by  $a \approx b$ .

The proposal of [6] takes into account preferences at the level of the attack relation. The idea is to look for the *critical attacks* and to *remove* them from the attack relation. An attack  $(a, b) \in \mathcal{R}$  is *critical* iff  $b > a$ . Formally this is done by defining a new binary relation, denoted by  $\text{Def}$ , as follows:  $a \text{ Def } b$  iff  $a\mathcal{R}b$  and not  $(b > a)$ . Thus, the relation  $\text{Def}$  contains all the attacks of  $\mathcal{R}$  except the critical ones (thus,  $\text{Def} \subseteq \mathcal{R}$ ). In order to evaluate the acceptability of the arguments of  $\mathcal{A}$ , Dung’s acceptability semantics are applied to the framework  $(\mathcal{A}, \text{Def})$ .

In [38], the preference relation  $\geq$  is given by arguments. The idea is that an argument may support a preference between two other arguments. Two attack relations are assumed: a classical one denoted by  $\mathcal{R}$ , and another relation,  $\mathcal{D}$ , that ranges from an argument of  $\mathcal{A}$  to an element of  $\mathcal{R}$ . An expression  $(a, (b, c))$  means that the argument  $a$  supports a preference of  $c$  over  $b$ . This preference conflicts with the fact that  $b$  attacks  $c$ . A new relation,  $\text{Def}$ , is defined as follows:  $a \text{ Def}_{\mathcal{S}} b$  iff  $a\mathcal{R}b$  and  $\nexists c \in \mathcal{S}$  such that  $(c, (a, b)) \in \mathcal{D}$ , where  $\mathcal{S} \subseteq \mathcal{A}$ . In [38], the author argues that his framework generalizes the one proposed in [6]. To put it differently, the framework proposed in the former work is grounded on the latter one.

The extension proposed in [17] assumes that a set  $\mathcal{V}$  of values is available. Each argument in  $\mathcal{A}$  promotes one value given by a function  $\text{val}$  (i.e.  $\text{val} : \mathcal{A} \mapsto \mathcal{V}$ ). Values may not have the same importance and this is captured by a binary relation  $\text{Pref}$ . This latter is assumed to be irreflexive, asymmetric and transitive. Like in [6], a new relation, called  $\text{defeats}$ , is defined as follows:  $a \text{ defeats } b$  iff  $a\mathcal{R}b$  and  $(\text{val}(b), \text{val}(a)) \notin \text{Pref}$ , thus  $\text{defeats} \subseteq \mathcal{R}$ . Dung’s acceptability semantics are applied to the framework  $(\mathcal{A}, \text{defeats})$ . In [34], a comparative study of the different approaches for taking into account preferences in argumentation frameworks has been carried out. This study reveals that the value-based framework is a particular instantiation of the framework proposed in [6] where the preference relation  $\geq$  in [6] comes from the importance of values.

In sum, the value-based framework is a particular case of the framework proposed in [6] which is in turn at the heart of the framework developed in [38]. Thus, if the framework developed in [6] suffers from any limit, then it is so for the two other works.

For simplicity reasons, we will use the framework proposed in [6] in order to illustrate the limits of existing approaches.

### 3 Critical examples

The three approaches [6, 17, 38] look for critical attacks, remove them from the attack relation, and then evaluate arguments on the basis of the remaining attacks. While this seems meaningful, removing attacks may lead to *conflicting extensions* in case of non-symmetric attack relations.

*Example 1* Assume that  $\mathcal{A} = \{a, b\}$  and  $\mathcal{R} = \{(a, b)\}$  ( $\mathcal{R}$  being not symmetric like undercut in [41]). Assume also that  $b > a$  (because, for instance, the value promoted by  $b$  is more important than the value promoted by  $a$ ). It can be checked that  $\text{Def} = \emptyset$  and  $\{a, b\}$  is a stable extension of the framework. Note that this extension is not conflict-free in the sense of  $\mathcal{R}$ . This is certainly in contradiction with the fact that an extension represents a coherent point of view.

In what follows, we show through a simple example that violating conflict-freeness is dangerous since it may lead to the violation of the rationality postulates proposed in [24].

Let  $\Sigma = \{x, \neg y, x \rightarrow y\}$  be a propositional knowledge base such that the formula  $x$  is *more certain* than the two other formulas. In what follows, an argument is defined as follows:

**Definition 3 (Argument)** Let  $\Sigma$  be a propositional knowledge base. An *argument* is a pair  $a = (H, h)$  s.t.

- $H \subseteq \Sigma$
- $H$  is consistent
- $H \vdash h$
- $\nexists H' \subset H$  such that  $H'$  is consistent and  $H' \vdash h$ .

The set  $H$  is the *support* of the argument and  $h$  is its *conclusion*. The following arguments are thus built from the base  $\Sigma$ .

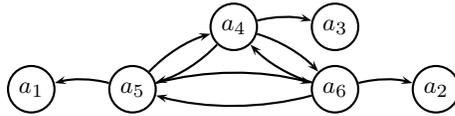
$a_1 : \langle \{x\}, x \rangle$	$a_2 : \langle \{\neg y\}, \neg y \rangle$
$a_3 : \langle \{x \rightarrow y\}, x \rightarrow y \rangle$	$a_4 : \langle \{x, \neg y\}, x \wedge \neg y \rangle$
$a_5 : \langle \{\neg y, x \rightarrow y\}, \neg x \rangle$	$a_6 : \langle \{x, x \rightarrow y\}, y \rangle$

The question now is which attack relation to choose? According to [1], the relation should not be symmetric since propositional logic is a particular case of Tarski's monotonic logics. Moreover, the knowledge base  $\Sigma$  contains a ternary minimal conflict (the three formulas of the base). Thus, we need an asymmetric one like *assumption attack* which has been proposed in [32] as follows:

**Definition 4 (Assumption attack)** An argument  $(H, h)$  *attacks* an argument  $(H', h')$ , denoted by  $(H, h) \mathcal{R}_{as} (H', h')$ , iff  $\exists h'' \in H'$  s.t.  $h \equiv \neg h''$ .

Moreover, it has been shown in [26] that the corresponding argumentation framework ensures sound results. Indeed, each stable extension of the framework returns a maximal consistent subbase of  $\Sigma$ . This result has been generalized to any Tarskian logic in [2].

The figure below depicts the attacks between the six above arguments with respect to assumption attack.



One may argue that other arguments (like  $\langle\{x\}, x \vee y\rangle$ ,  $\langle\{x\}, x \wedge x\rangle$ , ...) can be built from  $\Sigma$ , and thus the graph is bigger. This is certainly true. However, in a recent study [3], we have shown that when arguments are built from a propositional knowledge base and assumption attack is used, then the framework has a finite core of arguments which is sufficient to get the expected results. For instance, in the above example, it can be checked that the six arguments are sufficient to return the three maximal consistent subbases of  $\Sigma$  (of course when the priorities between the three formulas are ignored).

Assume now that the six arguments are compared using a relation which prefers strictly the argument  $a_1$  to all the other arguments and which equally prefers the five other arguments. An example of such a relation is the one based on the *weakest link principle* [18]. According to this principle, 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  $b$ . In this case, the base  $\Sigma$  is stratified into  $\Sigma_1 \cup \dots \cup \Sigma_n$  such that formulas of  $\Sigma_i$  have the same certainty level and are more certain than formulas in  $\Sigma_j$  where  $j > i$ . The stratification of  $\Sigma$  enables to define a certainty level of each subset  $\mathcal{S}$  of  $\Sigma$ . It is the highest number of stratum met by this subset. Formally:

$$\text{Level}(\mathcal{S}) = \max\{i \mid \exists x \in \mathcal{S} \cap \Sigma_i\} \text{ (with } \text{Level}(\emptyset) = 0\text{)}.$$

The above certainty level is used in [18] in order to define a total preorder on the set of arguments that can be built from a stratified knowledge base. The preorder is defined as follows:

**Definition 5 (Weakest link principle)** Let  $\Sigma = \Sigma_1 \cup \dots \cup \Sigma_n$  be a propositional knowledge base. An argument  $(H, h)$  is preferred to another argument  $(H', h')$ , denoted by  $(H, h) \geq_{WLP} (H', h')$ , iff  $\text{Level}(H) \leq \text{Level}(H')$ .

The three existing approaches [6, 17, 38] for preference-based argumentation remove the attack from  $a_5$  to  $a_1$  and obtain  $\mathcal{B} = \{a_1, a_2, a_3, a_5\}$  as a stable extension. Note that if we use the framework proposed in [38], then supplementary arguments, which specify that the argument  $a_1$  is stronger than the others, should be added. However, the framework will also compute the extension  $\mathcal{B}$ . This extension is not conflict-free with respect to the attack relation  $\mathcal{R}_{as}$ . Worse yet, it contains two arguments with contradictory conclusions ( $x$  and  $\neg x$ ). It is clear that this problem is due to the addition of preferences since as said before, when preferences are ignored the framework ensures sound results. What happens is that when an argument is stronger than its attacker, the attack is completely removed from the graph. By so doing an important information is lost. This information is the conflict that exists between the two arguments and consequently the two arguments may belong to the same extension. Note that this observation holds for any asymmetric relation and not only assumption attack. Thus an approach which removes attacks is not acceptable since it does not guarantee conflict-free extensions.

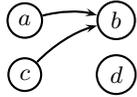
#### 4 A new approach for PAFs

The previous section has highlighted the limits of existing preference-based argumentation frameworks. Even if the idea pursued by these frameworks is meaningful, their results may violate the key property of conflict-freeness wrt the attack relation  $\mathcal{R}$ . This problem is mainly due to the fact of removing critical attacks from the framework.

In this section, we propose a new approach for preference-based argumentation which prevents the above problem. We consider a new line of research. Instead of changing the original attack relation, we take into account preferences when evaluating the arguments, i.e. at the semantics level. Our aim is not to define new acceptability semantics but to generalize the existing ones with preferences. Hence, when there is no critical attack, the extended semantics should return the same results as the basic ones (without preferences).

Our approach presents another novelty which consists of defining a semantics as a *dominance relation* on the powerset of the set  $\mathcal{A}$  of arguments. The best elements wrt this relation are the acceptable sets of arguments, i.e. the extensions. Recall that existing semantics divide the powerset of  $\mathcal{A}$  into two subsets: extensions and non-extensions. The former are better than the latter, but they do not say anything about non-extensions. However, in some applications, one may want to compare some sets of arguments. For instance, after a dialogue between two agents, an observer may want to compare the two sets of arguments exchanged by the two agents. Another important application in which the comparison of sets of arguments is crucial is decision making. Let us consider the following example.

*Example 2* Assume that  $\mathcal{A} = \{a, b, c, d\}$ ,  $a \approx c$ ,  $a > b$ ,  $d > a$  and let the attack relation be as depicted in the figure below.



Assume that both arguments  $a$  and  $c$  support an option  $o_1$ , the argument  $b$  supports an option  $o_2$  and the argument  $d$  supports an option  $o_3$ . Note that in this example there is no critical attack. Thus, the whole graph is considered for evaluating the arguments. According to Dung's semantics, the set  $\{a, c, d\}$  is the only stable extension of this framework, and it is thus preferred to all the other subsets of  $\mathcal{A}$ . This set contains arguments concerning the two options  $o_1$  and  $o_3$ . Thus, these latter are seen as equally preferred. However, it is clear from the preference relation  $\geq$  that the set  $\{d\}$  is preferred to  $\{a, c\}$ , thus option  $o_3$  is better than option  $o_1$ . Note also that it is not possible to conclude that the set  $\{a, c\}$  is preferred to the set  $\{b\}$ , and thus that option  $o_1$  is preferred to option  $o_2$ . It is neither possible to conclude that the set  $\{d\}$  is preferred to  $\{b\}$ , thus that option  $o_3$  is preferred to  $o_2$ .

Note that in [10], an argument-based decision framework has been proposed. It follows a two steps process: it computes first all the acceptable arguments (using Dung's semantics) then it applies a *decision principle* in order to rank-order options. As argued in [8], with such an approach the output of the argumentation framework is not yet

the output of the decision process. Thus, it is not possible to benefit from the whole powerful of argumentation theory (namely proof procedures). In [8], a one step approach where the argumentation framework returns directly the best option has been developed. However, in that paper, *only* one decision principle (the pessimistic qualitative criterion) has been encoded. Moreover, it returns only the best option. Thus, in case of Example 2, it is not possible for such a framework to conclude that option  $o_1$  is better than option  $o_2$ . This information is important in negotiation dialogues where agents make concessions (i.e. proposes less preferred options). It is also useful in any application where a complete ordering on a set of options is required.

In sum, we think that defining a semantics as a preference relation allows the comparison of any pair of subsets of arguments (on the basis of attacks and argument orderings). Such an approach is very suitable for decision making and solves the problems encountered by existing decision models [8, 10].

Before defining formally the new semantics, let us first introduce some notations and concepts. The set  $\mathcal{A}$  and the two relations  $\mathcal{R}$  and  $\succeq$  constitute a preference-based argumentation framework (PAF).

**Definition 6 (PAF)** A PAF is a tuple  $(\mathcal{A}, \mathcal{R}, \succeq)$  where  $\mathcal{A}$  is a set of arguments,  $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$  an attack relation, and  $\succeq \subseteq \mathcal{A} \times \mathcal{A}$  a (partial or total) preorder.

Assumptions: Without loss of generality, the following assumptions are made throughout the paper. For a PAF  $(\mathcal{A}, \mathcal{R}, \succeq)$ :

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

Notation: Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  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  $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ .

As already explained, a semantics for evaluating arguments of a PAF is defined as a binary relation on the powerset  $\mathcal{P}(\mathcal{A})$  of  $\mathcal{A}$ . Such a relation will be denoted by  $\succeq$ .

**Definition 7 (New semantics)** A *semantics* is a binary relation  $\succeq$  on the powerset  $\mathcal{P}(\mathcal{A})$  of  $\mathcal{A}$ . 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  $(\mathcal{E}' \succeq \mathcal{E})$ .

The maximal elements of such a relation are defined as follows.

**Definition 8 (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:

1.  $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A}), \mathcal{E} \succeq \mathcal{E}'$ ,
2. No strict superset of  $\mathcal{E}$  verifies (1).

Let  $\succeq_{max}$  denote the set of maximal sets wrt  $\succeq$ .

Like existing acceptability semantics, preference-based semantics should satisfy some basic requirements. Thus, not any relation  $\succeq$  can be used for evaluating arguments in a PAF. An appropriate relation should satisfy at least three postulates. Before presenting them, let us introduce a useful notation.

Notation: The writing  $\frac{X_1 \dots X_n}{Y}$  means that if  $X_1 \dots$  and  $X_n$  hold, then  $Y$  holds as well.

The first postulate states that any conflict-free set of arguments should be strictly preferred to any conflicting one. This is natural since a conflict-free set represents a coherent point of view while a conflicting set represents conflicting views.

**Postulate 1 (P1)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ .

$$\frac{\mathcal{E} \in \mathcal{CF}(\mathcal{T}) \quad \mathcal{E}' \notin \mathcal{CF}(\mathcal{T})}{\mathcal{E} \succ \mathcal{E}'}$$

Postulate P1 ensures conflict-freeness for the extensions of any PAF. Indeed, the best elements of any dominance relation satisfying this postulate are conflict-free.

*Property 1* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF. If a dominance relation  $\succeq$  satisfies postulate P1, then each element of the set  $\succeq_{max}$  is conflict-free wrt  $\mathcal{R}$ .

The second postulate describes the role of the attack relation. It shows that an attack should win when it is not critical. This is in some sense the basic idea behind all existing semantics in the literature.

**Postulate 2 (P2)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $a, a' \in \mathcal{A}$ .

$$\frac{a \mathcal{R} a' \quad \neg(a' \mathcal{R} a) \quad \neg(a' > a)}{\{a\} \succ \{a'}}$$

The third postulate ensures that preferences are privileged in critical attacks. This is in fact the idea defended in previous works on PAFs (e.g. [6,17]). Indeed, if an argument  $a$  attacks another argument  $a'$  and  $a' > a$ , then the set  $\{a'\}$  is privileged. Thus,  $\{a'\}$  should be strictly preferred to  $\{a\}$ . However, if the two arguments are equally preferred or incomparable or even  $a > a'$ , then the set  $\{a\}$  should be strictly preferred to  $\{a'\}$ .

**Postulate 3 (P3)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $a, a' \in \mathcal{A}$ .

$$\frac{a \mathcal{R} a' \quad a' > a}{\{a'\} \succ \{a}}$$

We are now ready to define what is a semantics for evaluating the arguments of a PAF. It is a binary relation (called also dominance relation) on the powerset of the set of arguments which satisfies the above postulates. The acceptable sets of arguments are the best elements of the dominance relation.

**Definition 9 (Semantics for PAFs)** An *acceptability semantics* for a PAF  $(\mathcal{A}, \mathcal{R}, \succeq)$  is defined by a dominance relation  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$  which satisfies postulates P1, P2 and P3. The *acceptable sets* of arguments (or extensions) of  $(\mathcal{A}, \mathcal{R}, \succeq)$  are the elements of  $\succeq_{max}$ .

## 5 Generalizing Dung's semantics with preferences

In this section, we propose three new semantics which generalize respectively stable, preferred and grounded semantics introduced in [29]. Before presenting them, let us first define formally how a semantics can generalize another one.

**Definition 10 (Generalizing a semantics)** A dominance relation  $\succeq$  *generalizes semantics*  $x$  iff for all PAF  $(\mathcal{A}, \mathcal{R}, \succeq)$ , if  $\nexists a, b \in \mathcal{A}$  such that  $a\mathcal{R}b$  and  $b > a$ , then  $\succeq_{max} = \text{Ext}_x(\mathcal{F})$  where  $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ .

Informally speaking, a dominance relation generalizes a given semantics iff its best elements are exactly the extensions of the basic framework (i.e. without preferences) wrt that semantics in case there are no critical attacks.

### 5.1 Generalizing stable semantics

In the previous section, we have shown that our new semantics are defined as dominance relations on the powerset of the set of arguments. 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 powerset 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  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ . The equality  $\text{Ext}(\mathcal{F}) = \succeq_{max}$  holds iff  $\forall \mathcal{E} \in \mathcal{P}(\mathcal{A})$ ,*

1. *if  $\mathcal{E} \notin \mathcal{CF}(\mathcal{F})$  then  $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$  s.t.  $\text{not}(\mathcal{E} \succeq \mathcal{E}')$ , and*
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  $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ ,  $\mathcal{E} \succeq \mathcal{E}'$ , and*
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.  $\text{not}(\mathcal{E} \succeq \mathcal{E}')$ .*

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 stable 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 3* Consider the argumentation framework depicted in the figure below.



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 3, 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. According to relation  $\succeq_1$ , the set  $\{\}$  is preferred to the set  $\{a, b\}$ .

In what follows, we present a new semantics, called *Pref-stable*, that generalizes stable semantics. This amounts to define a dominance relation which will be denoted

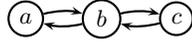
by  $\succeq_s$  and its best elements by  $\succeq_{s,max}$ . The idea behind this relation is the following: given two conflict-free sets of arguments,  $\mathcal{E}$  and  $\mathcal{E}'$ , we say that  $\mathcal{E}$  is better than  $\mathcal{E}'$  iff any argument in  $\mathcal{E}' \setminus \mathcal{E}$  is weaker than at least one argument in  $\mathcal{E} \setminus \mathcal{E}'$  or is attacked by it. Moreover, a conflict-free set of arguments is strictly preferred to a conflicting one, while conflicting sets are all incomparable. In fact, the relation  $\succeq_s$  extends the relation  $\succeq_1$  with preferences.

**Definition 11 (Pref-stable semantics)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . It holds that  $\mathcal{E} \succeq_s \mathcal{E}'$  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.  $(a\mathcal{R}a'$  and  $\text{not}(a' > a))$  or  $(a > a')$ .

Let us illustrate this definition through the following simple example.

*Example 4* Let  $\mathcal{A} = \{a, b, c\}$ ,  $a > b$  and  $\mathcal{R}$  is as depicted in the figure below:



The conflict-free sets of arguments are:  $\mathcal{E}_1 = \emptyset$ ,  $\mathcal{E}_2 = \{a\}$ ,  $\mathcal{E}_3 = \{b\}$ ,  $\mathcal{E}_4 = \{c\}$ , and  $\mathcal{E}_5 = \{a, c\}$ . It can be checked that the following relations hold:  $\mathcal{E}_2 \succeq_s \mathcal{E}_1$ ,  $\mathcal{E}_3 \succeq_s \mathcal{E}_1$ ,  $\mathcal{E}_4 \succeq_s \mathcal{E}_1$ ,  $\mathcal{E}_5 \succeq_s \mathcal{E}_1$ ,  $\mathcal{E}_5 \succeq_s \mathcal{E}_4$ ,  $\mathcal{E}_5 \succeq_s \mathcal{E}_2$ ,  $\mathcal{E}_5 \succeq_s \mathcal{E}_3$ ,  $\mathcal{E}_4 \succeq_s \mathcal{E}_3$ ,  $\mathcal{E}_3 \succeq_s \mathcal{E}_4$ , and  $\mathcal{E}_2 \succeq_s \mathcal{E}_3$ . It can also be checked that  $\succeq_{max} = \{\mathcal{E}_5\}$ .

The relation  $\succeq_s$  is in conformity with Definition 9. Indeed, it satisfies the three postulates  $P1$ ,  $P2$  and  $P3$ .

*Property 2* The relation  $\succeq_s$  satisfies postulates  $P1$ ,  $P2$  and  $P3$ .

Since the relation  $\succeq_s$  satisfies postulate  $P1$ , its extensions are conflict-free. The following result shows that they are even maximal (for set inclusion). Indeed, the relation  $\succeq_s$  privileges maximal sets.

*Property 3* Let  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . If  $\mathcal{E} \subsetneq \mathcal{E}'$  then  $\mathcal{E}' \succ_s \mathcal{E}$  (i.e.  $\mathcal{E}' \succeq_s \mathcal{E}$  and  $\text{not}(\mathcal{E} \succeq_s \mathcal{E}')$ ).

However, not any maximal conflict-free set of arguments is an extension (i.e. an element of  $\succeq_{s,max}$ ) as shown by the following example.

**Example 4 (Continued):** The set  $\mathcal{E}_3$  is maximal conflict-free but does not belong to  $\succeq_{s,max}$ .

From Property 3, it follows that Definition 8 can be simplified as follows:  $\mathcal{E} \in \succeq_{s,max}$  iff  $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A}), \mathcal{E} \succeq_s \mathcal{E}'$ . It is worth mentioning that this simplification is not possible for semantics that generalize preferred and grounded semantics. Finally, notice that the relation  $\succeq_s$  is *not transitive*. Indeed, in the previous example,  $\mathcal{E}_2 \succeq_s \mathcal{E}_3$  and  $\mathcal{E}_3 \succeq_s \mathcal{E}_4$  however, the two sets  $\mathcal{E}_2$  and  $\mathcal{E}_4$  are incomparable wrt  $\succeq_s$ .

The following theorem shows that Pref-stable semantics, i.e. the relation  $\succeq_s$ , generalizes stable semantics. Recall that this means that the two semantics coincide in case any attacked argument is not stronger than its attacker.

**Theorem 2** *The relation  $\succeq_s$  generalizes stable semantics.*

Finally, we can show that the proposed approach handles correctly the example discussed in Section 3. Namely, it can be checked that the corresponding PAF has exactly two extensions:  $\{\alpha_1, \alpha_2, \alpha_4\}$  (whose base is  $\{x, \neg y\}$ ) and  $\{\alpha_1, \alpha_3, \alpha_6\}$  (whose base is  $\{x, x \rightarrow y\}$ ), and that both of them are conflict-free and support consistent conclusions. This example will be treated in detail in Section 7.

## 5.2 Generalizing preferred semantics

In this section, we propose a new semantics, called *Pref-preferred*, that generalizes preferred semantics with preferences. It is defined by a dominance relation, denoted by  $\succeq_p$ . The basic idea behind this relation is that a set  $\mathcal{E}$  of arguments is better than another set  $\mathcal{E}'$  of arguments iff for every attack from  $\mathcal{E}'$  to  $\mathcal{E}$  which does not fail,  $\mathcal{E}$  is capable to defend the attacked argument and that for every attack from  $\mathcal{E}$  to  $\mathcal{E}'$  which fails, there is another attack from  $\mathcal{E}$  which defends the argument which failed in its attack.

**Definition 12 (Pref-preferred semantics)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . It holds that  $\mathcal{E} \succeq_p \mathcal{E}'$  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  $\forall a \in \mathcal{E}, \forall a' \in \mathcal{E}'$ , if  $(a' \mathcal{R} a$  and  $\text{not}(a > a')$ ) or  $(a \mathcal{R} a'$  and  $a' > a)$ , then  $\exists b \in \mathcal{E}$  such that  $(b \mathcal{R} a'$  and  $\text{not}(a' > b))$  or  $(a' \mathcal{R} b$  and  $b > a')$ .

Let us illustrate this definition through the next example.

**Example 4 (Continued):** It holds that  $\mathcal{E}_2 \succ_p \mathcal{E}_3$ ,  $\mathcal{E}_3 \succeq_p \mathcal{E}_4$ ,  $\mathcal{E}_4 \succeq_p \mathcal{E}_3$ ,  $\mathcal{E}_5 \succ_p \mathcal{E}_3, \dots$ . It can also be checked that  $\succeq_{p, \max} = \{\mathcal{E}_5\}$ .

Note that the relation  $\succeq_p$  is not transitive. However, it can be checked that it satisfies the three postulates *P1*, *P2* and *P3*. Thus, it encodes a semantics in the sense of Definition 9.

*Property 4* The relation  $\succeq_p$  satisfies postulates *P1*, *P2* and *P3*.

The above property ensures that the extensions of a PAF under Pref-preferred semantics are conflict-free. The following result shows that this semantics generalizes Dung's preferred semantics.

**Theorem 3** *The relation  $\succeq_p$  generalizes preferred semantics.*

In Dung's basic framework, every stable extension is a preferred one. We show that the same link holds in our setting. Namely, every pref-stable extension is a pref-preferred extension.

**Theorem 4** *For any PAF  $(\mathcal{A}, \mathcal{R}, \succeq)$ , it holds that  $\succeq_{s, \max} \subseteq \succeq_{p, \max}$ .*

### 5.3 Generalizing grounded semantics

In this section, we focus on grounded semantics and we generalize it with preferences. The new semantics is called *Pref-grounded* and is defined by a dominance relation which is denoted by  $\succeq_g$ . The basic idea behind this relation is that a set is not worse than another if it can strongly defend all its arguments against all attacks that come from the other set.

Before giving the formal definition of  $\succeq_g$ , let us first generalize the notion of strong defense by preferences. The idea is that an argument has either to be preferred to its attacker or has to be defended by arguments that themselves can be strongly defended without using the argument in question.

**Definition 13 (Strong defense)** Let  $\mathcal{E} \subseteq \mathcal{A}$ .  $\mathcal{E}$  *strongly defends* an argument  $a$  from attacks of a set  $\mathcal{E}'$ , denoted by  $sd(a, \mathcal{E}, \mathcal{E}')$ , iff  $\forall b \in \mathcal{E}'$  if  $(b\mathcal{R}a$  and  $\text{not}(a > b))$  or  $(a\mathcal{R}b$  and  $b > a)$ , then  $\exists c \in \mathcal{E} \setminus \{a\}$  such that  $((c\mathcal{R}b$  and  $\text{not}(b > c))$  or  $(b\mathcal{R}c$  and  $c > b)$  and  $sd(c, \mathcal{E} \setminus \{a\}, \mathcal{E}')$ .

If the third argument of  $sd$  is not specified, then  $sd(a, \mathcal{E}) \equiv sd(a, \mathcal{E}, \mathcal{A})$ .

Let us illustrate this notion through the following example.

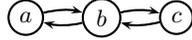
**Example 4 (Continued):**  $sd(a, \{a\}, \{b\})$  holds since  $a$  is strictly preferred to  $b$  thus it can defend itself. However, we have  $\neg sd(b, \{b\}, \{c\})$  since  $b$  cannot defend itself against  $c$ . On the other hand,  $sd(c, \{a, c\}, \{b\})$  holds since  $a$  can defend  $c$  against  $b$  and  $a$  is protected from  $b$  since it is strictly preferred to it.

The relation  $\succeq_g$  prefers subsets that strongly defend all their arguments. Namely,  $\mathcal{E} \succeq_g \mathcal{E}'$  iff  $\mathcal{E}$  strongly defends all its arguments against attacks from  $\mathcal{E}'$ .

**Definition 14 (Pref-grounded semantics)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}'$  be two subsets of  $\mathcal{A}$ . It holds that  $\mathcal{E} \succeq_g \mathcal{E}'$  iff:

- $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$  and  $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$ , or
- $\forall a \in \mathcal{E}$ , it holds that  $sd(a, \mathcal{E}, \mathcal{E}')$ .

*Example 5* Let  $\mathcal{A} = \{a, b, c\}$ ,  $b > a$ , and  $\mathcal{R}$  is as depicted in the figure below:



One can check that there is exactly one subset of  $\mathcal{A}$  which is preferred to all other subsets of arguments wrt  $\succeq_g$ . This set is the empty one. While we do have  $\{b\} \succeq_g \{a\}$ , we have  $\neg(\{b\} \succeq_g \{c\})$ , so  $\{b\}$  is not an extension of this PAF. We have also  $\neg(\{a\} \succeq_g \{b\})$ ,  $\neg(\{c\} \succeq_g \{b\})$  and  $\neg(\{a, c\} \succeq_g \{b\})$ . This is expected and a natural output since neither  $b$  nor  $c$  are capable to defend strongly themselves and, on the other hand, it can be said that  $a$  is the worst argument in this framework, thus not strong enough to be better than  $b$ .

The relation  $\succeq_g$  has exactly one best element, i.e. the set  $\succeq_{g, max}$  contains only one set of arguments. This is not surprising since Pref-grounded semantics intends to generalize the principle underlying Dung's grounded semantics.

*Property 5* The equality  $|\succeq_{g, max}| = 1$  holds.

The following result shows that the relation  $\succeq_g$  satisfies the three postulates  $P1$ ,  $P2$  and  $P3$ . Thus, its unique extension is conflict-free.

*Property 6* The relation  $\succeq_g$  satisfies postulates  $P1$ ,  $P2$  and  $P3$ .

Finally, the dominance relation  $\succeq_g$  generalizes grounded semantics.

**Theorem 5** *The relation  $\succeq_g$  generalizes grounded semantics.*

In Dung's basic framework, the grounded extension is a subset of the intersection of all preferred extensions. The same link exists between pref-grounded and pref-preferred extensions:

**Theorem 6** *For any PAF  $(\mathcal{A}, \mathcal{R}, \succeq)$ ,  $\succeq_{g,max} \subseteq \bigcap \mathcal{E}_i$  where  $\mathcal{E}_i \in \succeq_{p,max}$ .*

## 6 Characterizing Pref-stable semantics

In the previous section, we have proposed three particular semantics which generalize respectively stable, preferred and grounded semantics with preferences. What is worth mentioning is that the three corresponding dominance relations are *not unique*. There exist, for instance, other relations which may generalize stable semantics by preferences. Not surprisingly, the same is true for preferred and grounded semantics. This remark opens many new questions: How many dominance relations that generalize a given semantics do exist? Are some of them "better" than others? What are their properties? What are the differences between them? In the rest of the paper we focus on stable semantics and give a formal and precise answer to these questions.

### 6.1 Postulates

In this subsection, we characterize all the dominance relations  $\succeq$  that generalize stable semantics with preferences. For that purpose, we identify a set of postulates that such relations should satisfy. It is clear that the three postulates  $P1$ ,  $P2$  and  $P3$  are in that set. Postulate  $P1$  ensures 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. Postulates  $P2$  and  $P3$  describe when the attack relation should take precedence over the preference relation and when this latter is privileged. These two postulates are generalized by two other postulates in such a way to include the principle behind Dung's stable semantics. The first 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 that set), then  $\mathcal{E}$  cannot be preferred to  $\mathcal{E}'$ .

**Postulate 4 (P4)** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF, and  $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$  such that  $\mathcal{E} \cap \mathcal{E}' = \emptyset$ .*

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

The second postulate describes when a set is preferred to another. 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  $a$  wins a conflict against  $a'$ : either  $a$  attacks  $a'$  and  $a'$  does not defend itself since it is not stronger than  $a$  wrt  $\succeq$ , or  $a'$  attacks  $a$  but  $a$  is strictly preferred to  $a'$ .

**Postulate 5 (P5)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$  such that  $\mathcal{E} \cap \mathcal{E}' = \emptyset$ .

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

It can be checked that if a dominance relation satisfies the two postulates  $P4$  and  $P5$ , then it satisfies also  $P2$  and  $P3$ .

*Property 7* Let  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ . If  $\succeq$  satisfies postulates  $P4$  and  $P5$ , then it also satisfies postulates  $P2$  and  $P3$ .

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

**Postulate 6 (P6)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF, and  $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ . Then,

$$\frac{\mathcal{E} \succeq \mathcal{E}'}{\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}} \quad \text{and} \quad \frac{\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}}{\mathcal{E} \succeq \mathcal{E}'}$$

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

**Definition 15 (Pref-stable semantics)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF. A relation  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$  encodes pref-stable semantics iff it satisfies postulates  $P1, P4, P5$  and  $P6$ .

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 the relation  $\succeq_s$  given in Definition 11 is a pref-stable relation and satisfies the four postulates.

*Property 8*  $\succeq_s$  is a pref-stable relation.

Like stable semantics, there are several relations that encode pref-stable semantics. However, they all return the same pref-stable extensions.

**Theorem 7** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\succeq, \succeq' \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ . If  $\succeq$  and  $\succeq'$  are pref-stable relations, then  $\succeq_{max} = \succeq'_{max}$ .

Note that postulates  $P1, P4, P5$  and  $P6$  encode important properties of stable semantics enriched with preferences. However, it is worth noticing that no relation which generalizes stable semantics and verifies  $P1$  and  $P5$  is transitive. Note that this is not surprising since  $P5$  describes one of the basic properties of stable semantics, which is that a set attacking another one should win. This notion is not necessarily transitive since it is based on an attack relation which does not exhibit any property. Indeed, an attack relation is generally not a preorder.

*Property 9* There exists no transitive relation which generalizes stable semantics and satisfies postulates  $P1$  and  $P5$ .

Finally, we can show that a pref-stable semantics generalizes stable semantics. This means that when preferences 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 8** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF. Any pref-stable relation  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$  generalizes stable semantics.*

## 6.2 General and specific pref-stable relations

As already said, there are several relations that encode pref-stable semantics. The aim of this subsection is to define the upper and lower bounds of these relations. The most general pref-stable relation, denoted by  $\succeq_{gn}$ , returns  $\mathcal{E} \succeq_{gn} \mathcal{E}'$  if and only if it can be proved from the four postulates that  $\mathcal{E}$  must be preferred to  $\mathcal{E}'$ .

**Definition 16 (General pref-stable relation)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ .  $\mathcal{E} \succeq_{gn} \mathcal{E}'$  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  $\forall a' \in \mathcal{E}' \setminus \mathcal{E}, \exists a \in \mathcal{E} \setminus \mathcal{E}'$  such that  $(a \mathcal{R} a'$  and  $\text{not}(a' > a))$  or  $(a' \mathcal{R} a$  and  $a > a')$ .

*Property 10*  $\succeq_{gn}$  is a pref-stable relation.

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

**Definition 17 (Specific pref-stable relation)** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ .  $\mathcal{E} \succeq_{sp} \mathcal{E}'$  iff:

- $\mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$ , or
- $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$  and  $\forall a' \in \mathcal{E}' \setminus \mathcal{E}, \exists a \in \mathcal{E} \setminus \mathcal{E}'$  such that  $(a \mathcal{R} a'$  and  $\text{not}(a' > a))$  or  $(a > a')$ .

*Property 11*  $\succeq_{sp}$  is a pref-stable relation.

Let us illustrate the differences between the three particular relations  $\succeq_s$ ,  $\succeq_{sp}$  and  $\succeq_{gn}$  on the following example.

*Example 6* Let  $\mathcal{A} = \{a, b, c\}$ ,  $\mathcal{R} = \{(a, b)\}$  and  $a \geq c$ . For example, it holds that  $\{a\} \succeq_s \{c\}$ ,  $\{a\} \succeq_{sp} \{c\}$  and  $\neg(\{a\} \succeq_{gn} \{c\})$ . That is, for relations  $\succeq_s$  and  $\succeq_{sp}$  the strict preference between  $a$  and  $c$  is enough to prefer  $\{a\}$  to  $\{c\}$ . For relation  $\succeq_{gn}$ , 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  $\succeq_{sp}$ , all conflicting sets are equally preferred. For example,  $\{a, b, c\} \succeq_{sp} \{a, b\}$  and  $\{a, b\} \succeq_{sp} \{a, b, c\}$ . Besides, relations  $\succeq_s$  and  $\succeq_{gn}$  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  $\neg(\{a, b, c\} \succeq_s \{a, b\})$ .

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

**Theorem 9** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . Let  $\succeq$  be a pref-stable relation.*

- If  $\mathcal{E} \succeq_{gn} \mathcal{E}'$  then  $\mathcal{E} \succeq \mathcal{E}'$ .
- If  $\mathcal{E} \succeq \mathcal{E}'$  then  $\mathcal{E} \succeq_{sp} \mathcal{E}'$ .

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

### 6.3 Characterizing pref-stable extensions

As already said, the new approach for taking into account the strengths of arguments in an argumentation framework is sound and rich. It is sound since it guarantees conflict-free extensions, and it is rich since it provides more information than existing approaches. Indeed, not only it computes the acceptable sets of arguments, but it also compares the remaining ones. This comparison is of great importance in some applications like decision making and dialogues. However, it is less crucial in some other applications like handling inconsistency in knowledge bases. In this case, one looks only for the sets of arguments which support ‘good’ conclusions and does not bother about the other arguments. It is thus important to be able to characterize the extensions under a given semantics without comparing all the subsets of arguments, and thus without referring to pref-stable relations. The next theorem gives such a characterization.

**Theorem 10** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\succeq$  be a pref-stable relation.*

$\mathcal{E} \in \succeq_{max}$  iff:

- $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ , and
- $\forall a' \in \mathcal{A} \setminus \mathcal{E}, \exists a \in \mathcal{E}$  such that  $(a\mathcal{R}a'$  and  $\text{not}(a' > a))$  or  $(a'\mathcal{R}a$  and  $a > a')$ .

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  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$ . We compute an AF  $\mathcal{F} = (\mathcal{A}, \mathcal{R}')$  where  $\mathcal{R}'$  is defined as follows:

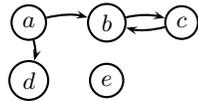
$$\begin{cases} \text{If } (a, b) \in \mathcal{R} \text{ and } \text{not}(b > a) \text{ then } (a, b) \in \mathcal{R}' \\ \text{If } (a, b) \in \mathcal{R} \text{ and } b > a \text{ then } (b, a) \in \mathcal{R}' \end{cases}$$

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

**Theorem 11** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\succeq$  be a pref-stable relation. Let  $\mathcal{R}' = \{(a, b) \mid a, b \in \mathcal{A}, (a\mathcal{R}b \text{ and } \text{not}(b > a)) \text{ or } (b\mathcal{R}a \text{ and } a > b)\}$ . It holds that  $\succeq_{max} = \text{Ext}((\mathcal{A}, \mathcal{R}'))$ .*

Let us illustrate this result through an example.

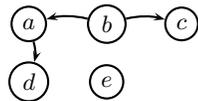
*Example 7* Let  $\mathcal{A} = \{a, b, c, d, e\}$  and  $\mathcal{R}$  is as depicted in figure below:



Assume that  $b > a$ ,  $b > c$  and  $e > d$ . Note that this framework has two critical attacks:  $(a, b)$  and  $(c, b)$ .

It can be checked that any pref-stable relation will return exactly one pref-stable extension:  $\succeq_{max} = \{\{b, d, e\}\}$ .

Let us now consider the following argumentation framework that is obtained after inverting the arrows of the two critical attacks.



It is easy to check that the only stable extension of this framework is the set  $\{b, d, e\}$ .

In [31], Dunne has investigated the computational properties of Dung's argumentation framework. He has shown that deciding whether a given set of arguments is a stable extension of AF can be performed in polynomial time. However, deciding whether an argumentation framework has a stable extension is NP-complete. The problems of deciding whether an argument is in some stable extension or in every stable extension are respectively NP-complete and coNP-complete. An important question is whether these results remain the same in our approach. The following result shows that the answer is yes. This is mainly due to the result of Theorem 11 which translates in a polynomial time a PAF into a basic argumentation framework.

**Theorem 12** *Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{E} \subseteq \mathcal{A}$ .*

- *Deciding whether  $\mathcal{E}$  is a pref-stable extension of  $\mathcal{T}$  is polynomial.*
- *Deciding whether the PAF  $\mathcal{T}$  has a pref-stable extension is NP-complete.*
- *Deciding whether an argument is in some pref-stable extension of  $\mathcal{T}$  is NP-complete.*
- *Deciding whether an argument is in every pref-stable extension of  $\mathcal{T}$  is coNP-complete.*

## 7 Computing preferred sub-theories with PAFs

An important problem in the management of knowledge-based systems is the handling of inconsistency. Classical logic has many appealing features for knowledge representation and reasoning, but unfortunately when reasoning with inconsistent information, i.e. drawing conclusions from an inconsistent knowledge base, the set of classical consequences is trivialized. To solve this problem, a *coherence-based* approach was initiated in [45]. It proposes to give up some formulas of the knowledge base in order to get one or several consistent subbases of the original base. Then plausible conclusions may be obtained by applying classical entailment on these subbases. In [26], it has been shown that the results of this approach can be recovered within Dung's argumentation framework [29]. Indeed, there is a full correspondence between the maximal consistent

subbases of a given inconsistent knowledge base and the stable extensions of the argumentation system built over the same base. In [45], the formulas of the knowledge base are assumed to be equally preferred. This assumption has been discarded in [21]. Indeed, a knowledge base is equipped with a total preorder. Thus, instead of computing the maximal consistent subbases, *preferred sub-theories* are computed. These sub-theories are consistent subbases that privilege the most important formulas.

In this section, we show that there is a full correspondence between the preferred sub-theories proposed in [21] and the pref-stable extensions of an instance of our PAF. The correspondence is obtained by choosing appropriately the main components of a PAF: the definition of an argument, the attack relation, and the preference relation between arguments.

### 7.1 Coherence-based approach for handling inconsistency

The coherence-based approach for handling inconsistency in a propositional knowledge base  $\Sigma$  follows two steps: At the first step, some subbases of  $\Sigma$  are chosen. In [45], these subbases are the maximal (for set inclusion) consistent ones. At the second step, an inference mechanism is chosen. This latter defines the inferences to be made from  $\Sigma$ . An example of inference mechanism is the one that infers a formula if it is a classical conclusion of all the chosen subbases. Several works have been done on choosing the subbases, in particular when  $\Sigma$  is equipped with a (total or partial) preorder  $\succeq$  ( $\succeq \subseteq \Sigma \times \Sigma$ ). Recall that when  $\succeq$  is total,  $\Sigma$  is stratified into  $\Sigma_1 \cup \dots \cup \Sigma_n$  s.t.  $\forall i, j$  with  $i \neq j$ ,  $\Sigma_i \cap \Sigma_j = \emptyset$ . Moreover,  $\Sigma_1$  contains the most important formulas while  $\Sigma_n$  contains the weakest ones.

In [21], the knowledge base  $\Sigma$  is equipped with a total preorder. The chosen subbases privilege the most important formulas.

**Definition 18 (Preferred sub-theory)** Let  $\Sigma$  be stratified into  $\Sigma_1 \cup \dots \cup \Sigma_n$ . A *preferred sub-theory* is a set  $\mathcal{S} = \mathcal{S}_1 \cup \dots \cup \mathcal{S}_n$  such that  $\forall k \in [1, n]$ ,  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_k$  is a maximal (for set inclusion) consistent subbase of  $\Sigma_1 \cup \dots \cup \Sigma_k$ .

*Example 8 (Example of Section 3 cont.)* The knowledge base  $\Sigma = \Sigma_1 \cup \Sigma_2$  with  $\Sigma_1 = \{x\}$  and  $\Sigma_2 = \{x \rightarrow y, \neg y\}$  has two preferred sub-theories:  $\mathcal{S}_1 = \{x, x \rightarrow y\}$  and  $\mathcal{S}_2 = \{x, \neg y\}$ .

It can be shown that the preferred sub-theories of a knowledge base  $\Sigma$  are maximal (wrt set inclusion) consistent subbases of  $\Sigma$ .

### 7.2 Computing sub-theories with a PAF

This section shows how an instance of our PAF computes the preferred sub-theories of a propositional knowledge base  $\Sigma$ .

**Assumption:** In the rest of this paper, we assume that a knowledge base  $\Sigma$  contains only consistent formulas.

Before presenting the instance, let us first introduce some useful notations.

Notations: Let  $a = (H, h)$  be an argument (in the sense of Definition 3). The functions **Supp** and **Conc** return respectively the support  $H$  and the conclusion  $h$  of the argument  $a$ . For  $\mathcal{S} \subseteq \Sigma$ ,  $\mathbf{Arg}(\mathcal{S}) = \{(H, h) \mid (H, h) \text{ is an argument in the sense of Definition 3 and } H \subseteq \mathcal{S}\}$ . Thus,  $\mathbf{Arg}(\Sigma)$  denotes the set of all the arguments that can be built from the whole knowledge base  $\Sigma$ . Finally, for  $\mathcal{E} \subseteq \mathbf{Arg}(\Sigma)$ ,  $\mathbf{Base}(\mathcal{E}) = \bigcup \mathbf{Supp}(a)$  where  $a \in \mathcal{E}$ .

The following result summarizes some useful properties of the two functions: **Arg** and **Base**.

*Property 12*

- For any consistent subbase  $\mathcal{S} \subseteq \Sigma$ ,  $\mathcal{S} = \mathbf{Base}(\mathbf{Arg}(\mathcal{S}))$ .
- For any  $\mathcal{E} \subseteq \mathbf{Arg}(\Sigma)$ ,  $\mathcal{E} \subseteq \mathbf{Arg}(\mathbf{Base}(\mathcal{E}))$ .
- A set  $\mathcal{S} \subseteq \Sigma$  is consistent *iff*  $\mathbf{Arg}(\mathcal{S})$  is conflict-free.

The instance uses all the arguments that can be built from  $\Sigma$  using Definition 3 (i.e. the set  $\mathbf{Arg}(\Sigma)$ ), the attack relation  $\mathcal{R}_{as}$  given in Definition 4, and the preference relation  $\geq_{WLP}$  introduced in Definition 5. Recall that the relation  $\geq_{WLP}$  is based on the weakest link principle and privileges the arguments whose less important formulas are more important than the less important formulas of the other arguments. The PAF that will be used is thus  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$ .

The first result shows that from a preferred sub-theory, it is possible to build a unique pref-stable extension of the PAF  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$ .

**Theorem 13** *Let  $\Sigma$  be a stratified knowledge base. For all preferred sub-theory  $\mathcal{S}$  of  $\Sigma$ , it holds that:*

- $\mathbf{Arg}(\mathcal{S})$  is a pref-stable extension of  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$
- $\mathcal{S} = \mathbf{Base}(\mathbf{Arg}(\mathcal{S}))$

Similarly, we show that each pref-stable extension of  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$  is built from a unique preferred sub-theory of  $\Sigma$ .

**Theorem 14** *Let  $\Sigma$  be a stratified knowledge base. For all pref-stable extension  $\mathcal{E}$  of  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$ , it holds that:*

- $\mathbf{Base}(\mathcal{E})$  is a preferred sub-theory of  $\Sigma$
- $\mathcal{E} = \mathbf{Arg}(\mathbf{Base}(\mathcal{E}))$

The next result shows that there exists a one-to-one correspondence between the preferred sub-theories of  $\Sigma$  and the pref-stable extensions of  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$ .

**Corollary 1** *Let  $\mathcal{T} = (\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$  be a PAF over a stratified knowledge base  $\Sigma$ . The pref-stable extensions of  $\mathcal{T}$  are exactly the  $\mathbf{Arg}(\mathcal{S})$  where  $\mathcal{S}$  ranges over the preferred sub-theories of  $\Sigma$ .*

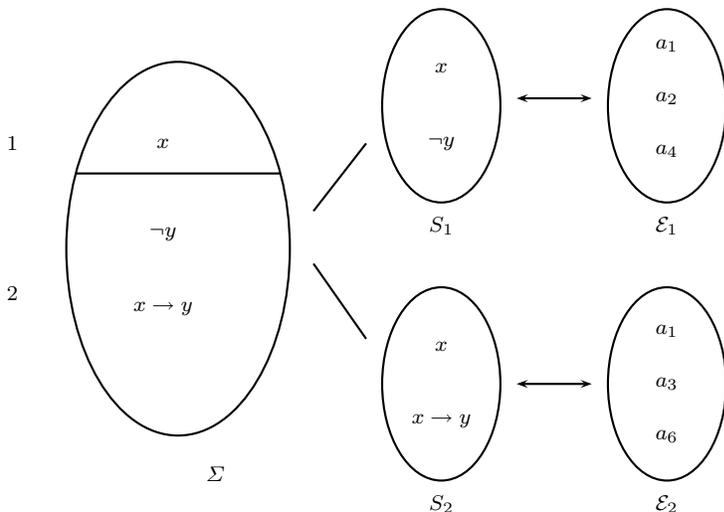
From the above result, it follows that the PAF  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$  has at least one pref-stable extension.

**Corollary 2** *The PAF  $(\mathbf{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$  has at least one pref-stable extension.*

*Example 9 (Example of Section 3 cont.)* Figure 1 shows the two preferred sub-theories of  $\Sigma$  as well as the two corresponding pref-stable extensions of the PAF.

The results of this section show that our approach is sound. It recovers the well-known preferred sub-theories.

**Fig. 1** Preferred sub-theories of  $\Sigma$  + Pref-stable extensions of  $(\text{Arg}(\Sigma), \mathcal{R}_{as}, \geq_{WLP})$



## 8 Related work

Preferences play a key role in nonmonotonic reasoning [23]. 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 [14], we have shown 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. We have also shown that the two roles of preferences are completely independent since none of them can be modeled by the other one.

The first role has largely been discussed in existing literature. It goes back to the paper [46]. In that work, the authors have defined an AF in which arguments are built from a propositional knowledge base. Arguments grounded on specific information are 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 [6] then in [17] to any AF and to any preference relation. Unfortunately, the approach followed in [6,17] 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 2 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. Our approach overcomes these limits.

Another work which handles correctly the problem of critical attacks is that proposed in [43, 39]. 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.

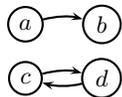
To the best of our knowledge, the only work on refinement (i.e. the second role of preferences) is that appeared in [28]. However, in that work the attack relation is assumed to be symmetric. Finally, we would like to mention the work done in [33]. In this paper, the author made a survey of the critics presented in [12, 28] 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 [28]. 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. In our paper, we have proposed a framework which handles correctly critical attacks. The results returned by this framework can then be refined by the relation proposed in [28].

Answer set programming is closely related to argumentation theory. Indeed, some correspondences can be made between the two theories. In [22], an axiomatic approach was proposed in order to take into account priorities between rules in an answer set program. Two principles were particularly introduced. If we translate them into an argumentation context, they would be defined as follows:

**Principle 1.** Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$  be a PAF. Let  $\mathcal{E}_1$  and  $\mathcal{E}_2$  be two extensions of the AF  $(\mathcal{A}, \mathcal{R})$  such that  $\mathcal{E}_1 = \mathcal{E} \cup \{a\}$  and  $\mathcal{E}_2 = \mathcal{E} \cup \{b\}$ . If  $a > b$ , then  $\mathcal{E}_1$  should be the only extension of  $\mathcal{T}$ .

This postulate is suitable for the second role of preferences and not for handling critical attacks as shown by the following example.

*Example 10* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$  be a PAF such that  $\mathcal{A} = \{a, b, c, d\}$ ,  $b > a$ ,  $c > d$  and  $\mathcal{R}$  is as depicted in the figure below.

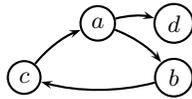


The AF  $(\mathcal{A}, \mathcal{R})$  has two stable extensions:  $\{a, c\}$  and  $\{a, d\}$ . According to the above postulate, the set  $\{a, c\}$  should be the only extension of the PAF  $\mathcal{T}$ . This is not realistic since the attack from  $a$  to  $b$  is critical. Thus, the argument  $b$  should win. According to our approach, the only extension of this PAF is the set  $\{b, c\}$ .

The second principle states that adding a rule which is not applicable in a preferred belief set can never render this belief set non preferred unless new preference information changes preferences among some of the old rules (e.g. via transitivity). In other

words, a belief set is not blamed for not applying rules which are not applicable. The following example shows again that this principle is not suitable in an argumentation context.

*Example 11* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$  be a PAF such that  $\mathcal{A} = \{a, b, c, d\}$ ,  $a \approx b \approx c$  and  $\mathcal{R}$  is as depicted in the figure below.



The only preferred extension of this PAF is the empty set ( $\emptyset$ ). If now we learn that  $d > a$ , thus the new PAF has a non-empty preferred extension which is  $\{d\}$ .

The two principles introduced in [22] allow to refine the results got in the flat case (without priorities) with priorities. However, a PAF does not necessarily refine the results of a basic argumentation framework. Thus, the principles that should be applied in both approaches should be different.

## 9 Conclusion

Several proposals have been made in the literature on integrating preferences in an argumentation framework. They argue that an attack may fail if the attacked argument is stronger than its attacker. They then suggest to remove such an attack from the argumentation framework and to evaluate the arguments on the basis of the remaining attacks. In this paper, we have shown that those proposals behave correctly when the attack relation is symmetric, however they may return unintended results when it is not. The reason is that by removing an attack, a crucial information is lost. Besides, a recent study ([1]) has shown that symmetric relations should be avoided when plugging Dung's approach upon a logical formalism. Indeed, if arguments are built from a logical knowledge base using a consequence operator which satisfies Tarski's axioms, then the framework with symmetric attack relation violates the rationality postulate on consistency. This latter is among the three basic properties that any argumentation framework should verify. Thus, a new approach for dealing with critical attacks is needed.

We have proposed here a novel approach which prevents the above limitations. Its idea is to take into account preferences at the semantics level, and to define thus *new* acceptability semantics that take into account both attacks and preferences between arguments. Existing semantics specify in a declarative way when a set of arguments is an extension (i.e. an acceptable set), but do not say anything on the subsets of arguments which are not extensions. However, in some applications like decision making, this information may be useful for comparing pairs of options. Thus, in this paper we have proposed a new line of research. A semantics is defined by a binary relation, called dominance relation, on the powerset of the set of arguments. The extensions of a PAF are the maximal elements of this relation. Thus, a semantics may compare any pair of subsets of arguments. Of course, not any binary relation can be used as a semantics. We have proposed three main postulates that any semantics should satisfy. The first one ensures the conflict-freeness of the extensions. It states that any

conflict-free set of arguments should be strictly preferred to a conflicting one. The two other postulates describe respectively the role of the attack relation and the role of the preference relation in a PAF. An attack should win if the attacked argument is not stronger than its attacker while a preference should win in case the attacked argument is stronger than its attacker. Another main contribution of our paper consists in generalizing Dung's acceptability semantics with preferences, namely preferred, stable and grounded semantics. The idea is to retrieve those semantics in case preferences do not conflict with the attacks. We have proposed a particular relation for each semantics, and we have shown that the links between Dung's semantics are preserved between their generalizations. We have also provided full characterizations of any relation which generalizes stable semantics as well as their extensions without referring to dominance relations. We have shown that integrating preferences does not change the worst-case complexity of different reasoning tasks under stable semantics. Finally, we have shown that our approach is valid since it allows to make a bridge between preference-based argumentation and coherence-based approach for handling inconsistency in weighted propositional knowledge bases. Namely, we have established full correspondences between the preferred sub-theories defined by Brewka in [21] and the generalized versions of stable extensions.

An extension of this work would be to characterize the different dominance relations that generalize preferred semantics and those which generalize grounded semantics. A similar work can be done on those semantics proposed in [16], ideal semantics [30] and semi-stable semantics [25]. Another future work consists of studying how the new semantics can be used in a decision making context in order to rank order a set of alternatives.

## Appendix

*Proof Property 1.* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF. Assume that  $\succeq$  is a dominance relation which satisfies postulate P1. Let us show that each element of the set  $\succeq_{max}$  is conflict-free wrt  $\mathcal{R}$ .

Assume that  $\mathcal{E} \in \succeq_{max}$ . Thus,  $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ ,  $\mathcal{E} \succeq \mathcal{E}'$ . In particular,  $\mathcal{E} \succeq \emptyset$ . Since  $\emptyset \in \mathcal{CF}(\mathcal{T})$ , then from Postulate P1,  $\mathcal{E} \in \mathcal{CF}(\mathcal{T})$ .

*Proof Theorem 1.* Let  $\mathcal{F} = (\mathcal{A}, \mathcal{R})$  be an AF and  $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ .

I. Assume that  $\mathbf{Ext}(\mathcal{F}) = \succeq_{max}$  and let us prove that the three conditions are satisfied.

1. Assume that  $\mathcal{E} \in \mathcal{P}(\mathcal{A})$  and  $\mathcal{E} \notin \mathcal{CF}(\mathcal{F})$ . So,  $\mathcal{E} \notin \mathbf{Ext}(\mathcal{F})$ , consequently,  $\mathcal{E} \notin \succeq_{max}$ . Thus,  $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$  s.t.  $\neg(\mathcal{E} \succeq \mathcal{E}')$ .
2. Assume that  $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$  and that  $\forall x' \notin \mathcal{E}$ ,  $\exists x \in \mathcal{E}$  s.t.  $x\mathcal{R}x'$ . Thus,  $\mathcal{E}$  is a stable extension of  $(\mathcal{A}, \mathcal{R})$ , and thus  $\mathcal{E} \in \succeq_{max}$ . Consequently,  $\forall \mathcal{E}' \in \mathcal{P}(\mathcal{A})$ ,  $\mathcal{E} \succeq \mathcal{E}'$ .
3. Assume that  $\mathcal{E} \in \mathcal{CF}(\mathcal{F})$  and  $\exists x' \in \mathcal{A} \setminus \mathcal{E}$  s.t.  $\nexists x \in \mathcal{E}$  and  $x\mathcal{R}x'$ . It is obvious that  $\mathcal{E}$  is not a stable extension of  $(\mathcal{A}, \mathcal{R})$ , thus  $\mathcal{E} \notin \mathbf{Ext}(\mathcal{F})$ . Since  $\mathbf{Ext}(\mathcal{F}) = \succeq_{max}$ , it follows that  $\mathcal{E} \notin \succeq_{max}$ . Thus,  $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$  s.t. not  $(\mathcal{E} \succeq \mathcal{E}')$ .

II. Assume that a relation  $\succeq$  satisfies the three conditions and let us prove that  $\mathbf{Ext}(\mathcal{F}) = \succeq_{max}$ .

- Let  $\mathcal{E}$  be a stable extension of  $(\mathcal{A}, \mathcal{R})$  and let  $\mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . From the second condition,  $\mathcal{E} \succeq \mathcal{E}'$ . Thus,  $\mathcal{E} \in \succeq_{max}$ .

- Assume that  $\mathcal{E} \in \succeq_{max}$ . Thus, for all  $\mathcal{E}' \in \mathcal{P}(\mathcal{A})$ ,  $\mathcal{E} \succeq \mathcal{E}'$ . From the first condition, it follows that  $\mathcal{E}$  is conflict-free. Assume that  $\mathcal{E} \notin \text{Ext}(\mathcal{F})$ . Thus,  $\exists x \notin \mathcal{E}$  and  $\mathcal{E}$  does not attack  $x$ . From the third condition,  $\exists \mathcal{E}' \in \mathcal{P}(\mathcal{A})$  s.t. not  $\mathcal{E} \succeq \mathcal{E}'$ . This contradicts the fact that  $\mathcal{E} \in \succeq_{max}$ .

*Proof Property 2.* Let us show that the relation  $\succeq_s$  satisfies postulates  $P1, P2$  and  $P3$ . From the first condition of Definition 11, it is clear that postulate  $P1$  is satisfied.

Let  $x$  and  $x'$  be two arguments. Since we assumed throughout the paper that there are no self-attacking arguments, then  $\{x\}$  and  $\{x'\}$  are conflict-free. Assume now that  $x\mathcal{R}x'$ ,  $\neg(x'\mathcal{R}x)$  and  $\neg(x' > x)$ . From the second condition of Definition 11, it follows that  $\{x\} \succ_s \{x'\}$ . Thus,  $\succeq_s$  satisfies postulate  $P2$ .

Assume now that  $x\mathcal{R}x'$  and  $x' > x$ . From the second condition of Definition 11, it follows that  $\{x\} \succeq_s \{x\}$ , however, not  $\{x\} \succeq_s \{x'\}$ . Thus,  $\{x'\} \succ_s \{x\}$ . Consequently, postulate  $P3$  is satisfied by  $\succeq_s$ .

*Proof Property 3.* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$  be a PAF and  $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ . Assume that  $\mathcal{E} \subsetneq \mathcal{E}'$ . Let us show that  $\mathcal{E}' \succ_s \mathcal{E}$ . Since  $\mathcal{E} \setminus \mathcal{E}' = \emptyset$ , then  $\mathcal{E}' \succeq_s \mathcal{E}$ . Let us now show that not  $(\mathcal{E} \succeq_s \mathcal{E}')$ . Since  $\mathcal{E} \subsetneq \mathcal{E}'$ , then  $\exists x' \in \mathcal{E}' \setminus \mathcal{E}$ . But, since  $\mathcal{E} \setminus \mathcal{E}' = \emptyset$ , we conclude that  $\nexists x \in \mathcal{E} \setminus \mathcal{E}'$  s.t.  $(x, x') \in >$  or  $((x, x') \in \mathcal{R} \wedge (x', x) \notin >)$ .

*Proof Theorem 2.* Let us show that the relation  $\succeq_s$  generalizes stable semantics.  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$  be a PAF. Assume that  $\nexists a, b \in \mathcal{A}$  s.t.  $a\mathcal{R}b$  and  $b > a$ . Assume that  $\mathcal{E}' \in \succeq_{s,max}$  and let us show that  $\mathcal{E}'$  is a stable extension of  $(\mathcal{A}, \mathcal{R})$ .

- Since  $\mathcal{E}' \in \succeq_{s,max}$  then it is conflict-free.
- We will now prove that  $\mathcal{E}'$  defends all its elements. Let us suppose that  $(\exists a \in \mathcal{E}') (\exists x \in \mathcal{A})$  s.t.  $(x, a) \in \mathcal{R} \wedge (\nexists y \in \mathcal{E}') (y, x) \in \mathcal{R}$ . Since  $\mathcal{E}'$  is conflict-free, then  $x \notin \mathcal{E}'$ . Let  $\mathcal{E} = \{x\} \cup \{t \in \mathcal{E}' \mid (x, t) \notin \mathcal{R} \wedge (t, x) \notin \mathcal{R}\}$ . It is clear the  $\mathcal{E}$  is conflict-free since  $\mathcal{E}$  is the union of two conflict-free sets which do not attack one another. Since  $\mathcal{E}' \in \succeq_{max}$  then  $\mathcal{E}' \succeq_s \mathcal{E}$ . In particular, since  $x \in \mathcal{E} \setminus \mathcal{E}'$ , then  $(\exists x' \in \mathcal{E}' \setminus \mathcal{E})$  s.t.  $((x', x) \in \mathcal{R} \wedge (x, x') \notin >) \vee (x', x) \in >$ . Since  $(\nexists y \in \mathcal{E}') (y, x) \in \mathcal{R}$ , then it must be the case that  $(x', x) \notin \mathcal{R}$  and  $(x', x) \in >$ . Since  $x' \in \mathcal{E}'$  and  $x' \notin \mathcal{E}$  then, with respect to definition of  $\mathcal{E}$ , from  $x' \notin \mathcal{E}$  we have that  $(x, x') \in \mathcal{R}$  or  $(x', x) \in \mathcal{R}$ . Since we have just seen that  $(x', x) \notin \mathcal{R}$ , it must be that  $(x, x') \in \mathcal{R}$ . Recall that we have  $(x', x) \in >$ . But we supposed that  $(\nexists z, z' \in \mathcal{A})$  s.t.  $(z, z') \in \mathcal{R}$  and  $(z', z) \in >$ . Contradiction. Thus,  $\mathcal{E}'$  defends its arguments.
- We have just shown that  $\mathcal{E}'$  is admissible, i.e., it is conflict-free and it defends all its arguments. We will now prove that  $\mathcal{E}'$  attacks all arguments in  $\mathcal{A} \setminus \mathcal{E}'$ . Let  $x \notin \mathcal{E}'$  be an argument and suppose that  $(\nexists y \in \mathcal{E}') (y, x) \in \mathcal{R}$ . Either  $x$  attacks some argument of  $\mathcal{E}'$  or not. If it is the case, i.e.,  $(\exists a \in \mathcal{E}')$  s.t.  $(x, a) \in \mathcal{R}$  then, since  $\mathcal{E}'$  defends all its elements, it holds that  $(\exists y \in \mathcal{E}')$  s.t.  $(y, x) \in \mathcal{R}$ . Contradiction. So, it must be that  $(\nexists a \in \mathcal{E}')$  s.t.  $(x, a) \in \mathcal{R}$ . This means that  $\mathcal{E} = \mathcal{E}' \cup \{x\}$  is conflict-free. According to Property 3, it holds that  $\neg(\mathcal{E}' \succeq_s \mathcal{E})$ . Contradiction with the fact that  $\mathcal{E}' \in \succeq_{s,max}$ .  
So,  $\mathcal{E}'$  is conflict-free and it attacks all arguments in  $\mathcal{A} \setminus \mathcal{E}'$ . This means that  $\mathcal{E}'$  is a stable extension of the framework  $(\mathcal{A}, \mathcal{R})$ .

Let  $\mathcal{E}'$  be a stable extension of the framework  $(\mathcal{A}, \mathcal{R})$  and let us prove that  $\mathcal{E}' \in \succeq_{s,max}$ .

- Since  $\mathcal{E}'$  is stable then it is conflict-free.

- We will prove that for an arbitrary conflict-free set of arguments  $\mathcal{E}$  it holds that  $\mathcal{E}' \succeq_s \mathcal{E}$ . Let  $\mathcal{E} \subseteq \mathcal{A}$  be a conflict-free set. If  $\mathcal{E} \setminus \mathcal{E}' = \emptyset$  the proof is over. If it is not the case, let  $x \in \mathcal{E} \setminus \mathcal{E}'$ . Since  $x \notin \mathcal{E}'$  and  $\mathcal{E}'$  is a stable extension, then  $(\exists x' \in \mathcal{E}') \text{ s.t. } (x', x) \in \mathcal{R}$ . We supposed that  $(\nexists z, z' \in \mathcal{A}) \text{ s.t. } (z, z') \in \mathcal{R} \text{ and } (z', z) \in >$ . Thus,  $(x, x') \notin >$ . Since  $x \in \mathcal{E} \setminus \mathcal{E}'$  was arbitrary, it holds that  $\mathcal{E}' \succeq_s \mathcal{E}$ .
- From Property 3, it follows that  $\mathcal{E}' \in \succeq_{s, max}$ .

*Proof Property 4.* Let us show that the relation  $\succeq_p$  satisfies postulates P1, P2 and P3. The definition of  $\succeq_p$  implies that P1 is ensured. Let us now suppose that for  $x, x' \in \mathcal{A}$  we have  $x\mathcal{R}x'$ ,  $\neg(x'\mathcal{R}x)$  and  $\neg(x' > x)$ . Since there are no self-attacking arguments, both  $\{x\}$  and  $\{x'\}$  are conflict-free. From Definition 12, we obtain  $\{x\} \succeq_p \{x'\}$  and  $\neg(\{x'\} \succeq_p \{x\})$ . Thus, P2 is verified. Let  $x\mathcal{R}x'$  and  $x' > x$ . From the same definition, this time we have that  $\neg(\{x\} \succeq_p \{x'\})$  and  $\{x'\} \succeq_p \{x\}$ . In other words,  $\{x'\} \succ_p \{x\}$ , which means that P3 is verified.

*Proof Theorem 3.* We will prove that preferred extensions of  $(\mathcal{A}, \mathcal{R})$  are exactly maximal elements of relation  $\succeq_p$ . Since we supposed that  $(\nexists x, y \in \mathcal{A}) \text{ s.t. } (x, y) \in \mathcal{R} \wedge (y, x) \in >$  then  $\mathcal{E}' \succeq_p \mathcal{E}$  iff  $(\forall x' \in \mathcal{E}') (\forall x \in \mathcal{E})$  if  $(x, x') \in \mathcal{R}$  then  $(\exists y' \in \mathcal{E}') \text{ s.t. } (y, x) \in \mathcal{R}$ .

$\Leftarrow$  Let  $\mathcal{E}'$  be a preferred extension of  $(\mathcal{A}, \mathcal{R})$ .

- Since  $\mathcal{E}'$  is a preferred extension then it is conflict-free.
- Let us prove that  $\mathcal{E}' \in \succeq_{p, max}$ . Suppose the contrary. This means that one of the following is true:
  1.  $(\exists \mathcal{E} \subseteq \mathcal{A}) \text{ s.t. } \mathcal{E} \text{ is conflict-free and } \neg(\mathcal{E}' \succeq_p \mathcal{E})$
  2.  $(\exists \mathcal{E} \subseteq \mathcal{A}) \text{ s.t. } \mathcal{E} \text{ is conflict-free } \wedge \mathcal{E}' \subsetneq \mathcal{E} \wedge (\forall \mathcal{E}'' \subseteq \mathcal{A}) \mathcal{E} \succeq_p \mathcal{E}''$
 Let (1) be the case. Since  $\neg(\mathcal{E}' \succeq_p \mathcal{E})$  then  $(\exists x' \in \mathcal{E}') (\exists x \in \mathcal{E}) \text{ s.t. } (x, x') \in \mathcal{R} \wedge (\nexists y' \in \mathcal{E}') \text{ s.t. } (y', x) \in \mathcal{R}$ . This leads to the conclusion that  $\mathcal{E}'$  does not defend its arguments, thus it cannot be a preferred extension. Contradiction. So, it must be that (2) holds. Since  $\mathcal{E}'$  is preferred and  $\mathcal{E}' \subsetneq \mathcal{E}$  then  $\mathcal{E}$  is not admissible. From the fact that  $\mathcal{E}$  is conflict-free, one concludes that it does not defend its arguments. Thus,  $(\exists x'' \in \mathcal{E}'' \setminus \mathcal{E}') \text{ s.t. } (\exists y \in \mathcal{A}) \text{ s.t. } (y, x'') \in \mathcal{R} \wedge (\nexists z'' \in \mathcal{E}'') \text{ s.t. } (z'', y) \in \mathcal{R}$ . Hence,  $\neg(\mathcal{E}'' \succeq_p \{y\})$ . Contradiction.

Let  $\mathcal{E}' \in \succeq_{p, max}$ . We will prove that  $\mathcal{E}'$  is a preferred extension of Dung's argumentation framework  $(\mathcal{A}, \mathcal{R})$ .

- Since  $\mathcal{E}' \in \succeq_{p, max}$  then it is conflict-free.
- Let us prove that  $\mathcal{E}'$  defends all its arguments. Suppose not. This means that  $(\exists y \in \mathcal{A}) \text{ s.t. } (y, x') \in \mathcal{R} \wedge (\nexists z' \in \mathcal{E}') \text{ s.t. } (z', y) \in \mathcal{R}$ . This means that  $\neg(\mathcal{E}' \succeq_p \{y\})$ . Contradiction.
- We have just seen that  $\mathcal{E}'$  is admissible. Let us prove that  $\mathcal{E}'$  is a preferred extension of  $(\mathcal{A}, \mathcal{R})$ . Suppose the contrary, i.e.,  $(\exists \mathcal{E} \subseteq \mathcal{A}) \text{ s.t. } \mathcal{E} \text{ is a preferred extension and } \mathcal{E}' \subsetneq \mathcal{E}$ . Since  $\mathcal{E}' \in \succeq_{p, max}$  then  $\mathcal{E} \notin \succeq_{p, max}$ . On the other hand, since  $\mathcal{E}$  is a preferred extension, then  $\mathcal{E} \in \succeq_{p, max}$ , as we have proved in the first part of this theorem. Contradiction.

*Proof Theorem 4.* We will prove that for any  $(\mathcal{A}, \mathcal{R}, \geq)$ , every pref-stable extension of this system is a pref-preferred extension of that system. In order to simplify the notation, we will write  $x\mathcal{D}y$  instead of  $(x\mathcal{R}y \text{ and not } (y > x))$  or  $(y\mathcal{R}x \text{ and } x > y)$ . It has been proved in [13] that if  $\mathcal{E} \in \succeq_{s, max}$  then  $\mathcal{E}$  is a Dung's stable extension of

the system  $(\mathcal{A}, \mathcal{D})$ . Then, from [29], we obtain that  $\mathcal{E}$  is Dung's preferred extension of the system  $(\mathcal{A}, \mathcal{D})$ . Now, we only have to prove that  $\mathcal{E} \in \succeq_{p, max}$ . It is obvious that  $\mathcal{E} \in \mathcal{CF}$ . Let  $\mathcal{E}' \subseteq \mathcal{A}$ . One can easily see that  $\mathcal{E} \succeq_p \mathcal{E}'$ ; this is a direct consequence of Definition 12. Let us prove that  $\nexists \mathcal{E}''$  s.t.  $\mathcal{E} \subsetneq \mathcal{E}''$  and  $\forall \mathcal{E}', \mathcal{E}'' \succeq_p \mathcal{E}'$ . Suppose the contrary; this would mean that  $\mathcal{E}''$  is admissible in  $(\mathcal{A}, \mathcal{D})$  which contradicts the fact that  $\mathcal{E}$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$ . Thus, it must be that  $\mathcal{E} \in \succeq_{p, max}$ , which ends the proof.

*Proof Property 5.* In the proof of this property, we will use Properties 50 and 51 from [15] which imply that for any argumentation system  $(\mathcal{A}, \mathcal{R})$ , for any  $x \in \mathcal{A}$ , we have that  $x \in \mathbf{GE}$  iff  $sd'(x, \mathbf{GE})$ , where  $\mathbf{GE}$  is the standard notation for grounded extension which will be used throughout the proof and  $sd'$  is the notion of strong defense as defined in Definition 13 in [15]. Note that for any  $a \in \mathcal{A}$ , for any  $\mathcal{A} \subseteq \mathcal{E}$ , we have that  $sd(a, \mathcal{E})$  iff  $(\forall b \in \mathcal{A}$  if  $b \mathcal{D} a$  then  $\exists c \in \mathcal{E} \setminus \{a\}$  s.t.  $c \mathcal{D} b$  and  $sd(c, \mathcal{E} \setminus \{a\})$ ), where we use  $x \mathcal{D} y$  as abbreviation for  $(x \mathcal{R} y$  and not  $(y > x)$ ) or  $(y \mathcal{R} x$  and  $x > y)$ . This proof will be based on the fact that we have  $sd(a, \mathcal{E})$  in  $(\mathcal{A}, \mathcal{R}, \geq)$  if and only if we have  $sd'(a, \mathcal{E})$  in  $(\mathcal{A}, \mathcal{D})$ . Thus, when we write  $sd(a, \mathcal{E})$ , we refer to system  $(\mathcal{A}, \mathcal{R}, \geq)$ , and when we use the function  $sd'$  and write  $sd'(a, \mathcal{E})$ , we refer to the corresponding system  $(\mathcal{A}, \mathcal{D})$ . By using this equivalence, we will prove that any set  $\mathcal{E} \subseteq \mathcal{A}$  is a pref-grounded extension of  $(\mathcal{A}, \mathcal{R}, \geq)$  iff  $\mathcal{E}$  is the grounded extension of  $(\mathcal{A}, \mathcal{D})$ .

$\Leftarrow$  Let  $\mathcal{E}$  be the grounded extension of  $(\mathcal{A}, \mathcal{D})$ . It is obvious that  $\mathcal{E} \in \mathcal{CF}$ . Let  $\mathcal{E}' \subseteq \mathcal{A}$ . Since  $\mathcal{E}$  is a grounded extension of  $(\mathcal{A}, \mathcal{D})$ , then from [15], we have  $x \in \mathcal{E} \Rightarrow sd'(x, \mathcal{E})$ . This means that we have  $sd(x, \mathcal{E})$  in  $(\mathcal{A}, \mathcal{R}, \geq)$ . Thus,  $sd(x, \mathcal{E}, \mathcal{E}')$  for any  $\mathcal{E}'$ , which means that  $\forall \mathcal{E}', \mathcal{E} \succeq_g \mathcal{E}'$ . Let us prove that  $\nexists \mathcal{E}'$  s.t.  $\mathcal{E}' \in \mathcal{CF}$  and  $\mathcal{E} \subsetneq \mathcal{E}'$  and  $\forall \mathcal{E}'', \mathcal{E}' \succeq_g \mathcal{E}''$ . Suppose the contrary. Suppose also that  $\forall x \in \mathcal{E}', sd(x, \mathcal{E}')$ . This means that  $\forall x \in \mathcal{E}', sd'(x, \mathcal{E}')$  in  $(\mathcal{A}, \mathcal{D})$ . Thus, from Proposition 51 of [15],  $\mathcal{E}' \subseteq \mathcal{E}$ , since  $\mathcal{E}$  is the grounded extension of  $(\mathcal{A}, \mathcal{D})$ . Contradiction, so it must be that  $\exists x \in \mathcal{E}'$  s.t.  $\neg sd(x, \mathcal{E}')$ . Thus,  $\exists y \in \mathcal{A}$  s.t.  $\neg(\mathcal{E} \succeq_g \{y\})$ . Contradiction, so we proved that  $\mathcal{E} \in \succeq_{g, max}$ .

$\Rightarrow$  Let  $\mathcal{E} \in \succeq_{g, max}$ . It is clear that  $\forall x \in \mathcal{E}, sd(x, \mathcal{E})$  in  $(\mathcal{A}, \mathcal{R}, \geq)$ . Thus,  $\forall x \in \mathcal{E}, sd'(x, \mathcal{E})$  in  $(\mathcal{A}, \mathcal{D})$ . From Proposition 51 of [15] we obtain  $\mathcal{E} \subseteq \mathbf{GE}$ , where  $\mathbf{GE}$  is the grounded extension of  $(\mathcal{A}, \mathcal{D})$ . Let us suppose that  $\mathcal{E} \subsetneq \mathbf{GE}$ . In the first part of this proof, we have shown that the grounded extension of  $(\mathcal{A}, \mathcal{D})$  is in  $\succeq_{g, max}$ . Contradiction, since we have supposed that  $\mathcal{E} \in \succeq_{g, max}$  and we have  $\mathcal{E} \subsetneq \mathbf{GE}$ . Thus,  $\mathcal{E} = \mathbf{GE}$ .

This shows that  $\mathcal{E} \in \succeq_{g, max}$  iff  $\mathcal{E}$  is the grounded extensions of the system  $(\mathcal{A}, \mathcal{D})$ . Since it has been shown in [29] that every argumentation system (without preferences) has exactly one grounded extension, we conclude that  $\succeq_{g, max}$  has exactly one element.

*Proof Property 6.* It is easy to see that P1 is satisfied. Let  $x \mathcal{R} x', \neg(x' \mathcal{R} x)$  and  $\neg(x' > x)$ . From definition of pref-grounded semantics, we have that  $\{x\} \succeq_g \{x'\}$  since  $sd(x, \{x\}, \{x'\})$ . On the other hand, the fact that  $\neg sd(x', \{x'\}, \{x\})$  implies that  $\neg(\{x'\} \succeq_g \{x\})$ . Thus, P2 is verified. Let us now prove that  $\succeq_g$  verifies P3. Let  $x \mathcal{R} x'$  a,d  $x' > x$ . In this case, we obtain  $\neg sd(x, \{x\}, \{x'\})$  and  $sd(x', \{x'\}, \{x\})$ , which means that  $\{x'\} \succ_g \{x\}$ .

*Proof Theorem 5.* We show that the grounded extension of  $(\mathcal{A}, \mathcal{R})$  is the only maximal element wrt  $\succeq_g$ . Since we supposed that  $(\nexists x, y \in \mathcal{A})$  s.t.  $(x, y) \in \mathcal{R} \wedge (y, x) \in >$  then we can simplify Definition 13 which becomes:  $sd(x, \mathcal{E}', \mathcal{E})$  iff  $(\forall y \in \mathcal{E})$  (if  $(y, x) \in \mathcal{R}$

then  $(\exists z \in \mathcal{E}' \setminus \{x\})$  s.t.  $((z, y) \in \mathcal{R} \wedge sd(z, \mathcal{E}' \setminus \{x\}, \mathcal{E}))$ . In this particular case when no attacked argument is strictly preferred to its attacker, our definition of  $sd(x, \mathcal{E})$  becomes exactly the same as Definition 13 in [15]. Thus, using Proposition 50 and Proposition 51 of the same paper, we conclude that  $x \in \mathbf{GE}$  iff  $sd(x, \mathbf{GE})$ , where  $\mathbf{GE}$  is the grounded extension of the framework  $(\mathcal{A}, \mathcal{R})$ .

$\Leftarrow$  Let  $\mathcal{E}'$  be the grounded extension of  $(\mathcal{A}, \mathcal{R})$ .

- Since  $\mathcal{E}'$  is the grounded extension then it is conflict-free.
- We will prove that for an arbitrary conflict-free set  $\mathcal{E} \subseteq \mathcal{A}$  it holds that  $\mathcal{E}' \succeq_g \mathcal{E}$ . Let  $\mathcal{E} \subseteq \mathcal{A}$  be conflict-free. Since  $\mathcal{E}'$  is the grounded extension then  $x \in \mathcal{E}' \Rightarrow sd(x, \mathcal{E}')$ . On the other hand,  $(\forall x \in \mathcal{E}') sd(x, \mathcal{E}')$  implies that  $sd(x, \mathcal{E}', \mathcal{E})$ . Thus,  $\mathcal{E}' \succeq_g \mathcal{E}$ . Since  $\mathcal{E}$  was arbitrary, then  $(\forall \mathcal{E} \subseteq \mathcal{A}) ((\mathcal{E} \text{ is conflict-free}) \Rightarrow (\mathcal{E}' \succeq_g \mathcal{E}))$ .
- We will now prove that  $(\nexists \mathcal{E} \subseteq \mathcal{A})$  s.t.  $\mathcal{E}$  is conflict-free and  $\mathcal{E}' \subsetneq \mathcal{E}$  and  $(\forall \mathcal{E}'' \subseteq \mathcal{A}) (\mathcal{E} \succeq_g \mathcal{E}'')$ . Suppose the contrary. Suppose also that  $(\forall x \in \mathcal{E}) sd(x, \mathcal{E})$ . If this is the case, according to Proposition 51 in [15],  $\mathcal{E} \subseteq \mathbf{GE}$ . Contradiction. So, it must be that  $(\exists x \in \mathcal{E})$  s.t.  $\neg sd(x, \mathcal{E})$ . Thus,  $(\exists y \in \mathcal{A})$  s.t.  $\neg sd(x, \mathcal{E}, \{y\})$ . Consequently,  $\neg(\mathcal{E} \succeq_g \{y\})$ . Contradiction. So, we have proved that  $\mathcal{E}' \in \succeq_{g, max}$ .

Let  $\mathcal{E}' \in \succeq_{g, max}$  and let us prove that  $\mathcal{E}' = \mathbf{GE}$ . Since  $(\forall x \in \mathcal{A}) \mathcal{E}' \succeq_g \{x\}$  then  $(\forall x' \in \mathcal{E}') sd(x', \mathcal{E}')$ . From the fact that  $(\forall x' \in \mathcal{E}') sd(x', \mathcal{E}')$  and Proposition 51 of [15] we have that  $\mathcal{E}' \subseteq \mathbf{GE}$ . Let us now prove that  $\mathcal{E}' = \mathbf{GE}$ . Suppose not, i.e., suppose that  $\mathcal{E}' \subsetneq \mathbf{GE}$ . We have proved in the first part of this theorem that  $\mathbf{GE} \in \succeq_{g, max}$ . Contradiction, since we have supposed that  $\mathcal{E}' \in \succeq_{g, max}$  and we have  $\mathcal{E}' \subsetneq \mathbf{GE}$ .

*Proof Theorem 6.* Let us suppose that  $\mathcal{E}$  is the pref-grounded extension of  $(\mathcal{A}, \mathcal{R}, \succeq)$ . By using the same reasoning as in the proof of Property 5, we conclude that  $\mathcal{E}$  is the grounded extension of the system  $(\mathcal{A}, \mathcal{D})$ , where  $x\mathcal{D}y$  is defined as  $(x\mathcal{R}y$  and not  $(y > x))$  or  $(y\mathcal{R}x$  and  $x > y)$ . In [29], it has been shown that the grounded extension of any argumentation system is a subset of the intersection of all preferred extensions of that system. Thus, in order to prove this property, it is sufficient to show that  $\forall \mathcal{E}$ , if  $\mathcal{E} \in \succeq_{g, max}$ , then  $\mathcal{E}$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$ , since this will imply that the intersection of preferred extensions of  $(\mathcal{A}, \mathcal{D})$  is a subset of the intersection of pref-preferred extensions of  $(\mathcal{A}, \mathcal{R}, \succeq)$ .

Let  $\mathcal{E} \in \succeq_{g, max}$ . Obviously,  $\mathcal{E} \in \mathcal{CF}$ . Let us prove that  $\mathcal{E}$  is admissible in  $(\mathcal{A}, \mathcal{D})$ . Let  $a \in \mathcal{E}, a' \notin \mathcal{E}$  and  $a'\mathcal{D}a$ . Since we supposed that  $\mathcal{E} \in \succeq_{g, max}$ , then  $\mathcal{E} \succeq_p \{a'\}$ . Consequently,  $\exists b \in \mathcal{E}$  s.t.  $b\mathcal{D}a'$ , so  $\mathcal{E}$  is admissible in  $(\mathcal{A}, \mathcal{D})$ . Let us suppose that  $\exists \mathcal{E}' \subseteq \mathcal{A}$ , s.t.  $\mathcal{E} \subsetneq \mathcal{E}'$  and  $\mathcal{E}'$  is admissible in  $(\mathcal{A}, \mathcal{D})$ . Then,  $\forall \mathcal{E}''$ , we have  $\mathcal{E}' \succeq_p \mathcal{E}''$ . Consequently, from the second item of Definition 8, we have that  $\mathcal{E} \notin \succeq_{g, max}$ , contradiction. Thus, it must be that  $\mathcal{E}$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$ . Since every pref-preferred extension of  $(\mathcal{A}, \mathcal{R}, \succeq)$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$ , then

$\bigcap_{\mathcal{E}_i \in \succeq_{p, max}} \mathcal{E}_i \subseteq \bigcap_{\mathcal{E}_j \text{ is a preferred extension of } (\mathcal{A}, \mathcal{D})} \mathcal{E}_j$ , which ends the proof of this property.

*Proof Property 7.* Let  $x, x' \in \mathcal{A}$ . Since there are no self-attacking arguments, then  $\{x\}, \{x'\} \in \mathcal{CF}$ . Let  $x\mathcal{R}x'$ ,  $\neg(x'\mathcal{R}x)$  and  $\neg(x' > x)$ . From the first part of Postulate 5 we have that  $\{x\} \succeq \{x'\}$ . From Postulate 4, we have  $\neg(\{x'\} \succeq \{x\})$ . Thus, Postulate 2 is verified. Let  $x\mathcal{R}x'$  and  $x' > x$ . From Postulate 5,  $\{x'\} \succeq \{x\}$ . Furthermore, Postulate 4 implies  $\neg(\{x\} \succeq \{x'\})$ . In sum,  $\{x'\} \succ \{x\}$ , which means that Postulate 3 is verified.

*Proof Property 8.* To show that  $\succeq_s$  is a pref-stable relation, we show that it satisfies postulates P4, P5, P6. Postulate 6 is satisfied since from the second item of the same

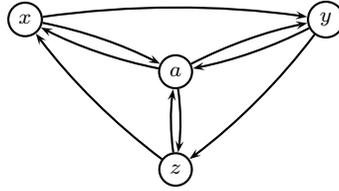
definition, when comparing two sets  $\mathcal{E}$  and  $\mathcal{E}'$ , common elements are not taken into account. The second condition of the definition of  $\succeq_s$  is exactly the negation of the condition of Postulate 4. Since Postulate 5 implies the second item of this definition, then it is verified.

*Proof Theorem 7.* We prove that all pref-stable relations return the same set of extensions.

$\Rightarrow$  Let  $\mathcal{E} \in \succeq_{max}$ . We will prove that  $\mathcal{E} \succeq'_{max}$ . From Postulate 1,  $\mathcal{E} \in \mathcal{CF}$ . Let  $\mathcal{E}' \subseteq \mathcal{A}$ . If  $\mathcal{E}'$  is not conflict-free then, from Postulate 1,  $\mathcal{E} \not\succeq' \mathcal{E}'$ . Else, from Postulate 6,  $\mathcal{E} \succeq' \mathcal{E}'$  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 disjoint conflict-free sets. If condition of Postulate 5 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 4 is not satisfied since  $\mathcal{E} \in \succeq_{max}$ . Thus, it must be that  $(\exists x' \in \mathcal{E}_2)$  s.t.  $(\nexists x \in \mathcal{E}_1)((x, x') \in \mathcal{R} \wedge (x', x) \notin \triangleright) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \triangleright)$  and  $(\exists x \in \mathcal{E}_1)(x, x') \in \triangleright$ . Let  $X = \{x \in \mathcal{E}_1 \mid (x, x') \in \triangleright\}$ .  $X$  is conflict-free. From Postulate 4,  $\neg(\mathcal{E}_1 \setminus X \succeq \{x'\})$ . Postulate 6 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  $\mathcal{E} \in \succeq_{max}$ . Thus, condition of Postulate 5 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  $\mathcal{E} \in \succeq'_{max}$ .

$\Leftarrow$  In the first part of proof, we showed that for all pref-stable relations  $\succeq_1, \succeq_2$ , it holds that if  $\mathcal{E} \in \succeq_{max}^1$  then  $\mathcal{E} \in \succeq_{max}^2$ . Contraposition of this rule gives specifies that if  $\mathcal{E} \notin \succeq_{max}^2$  then  $\mathcal{E} \notin \succeq_{max}^1$ . Since this was proved for arbitrary relations which satisfy  $P1, P4, P5$  and  $P6$ , we conclude: if  $\mathcal{E} \notin \succeq'_{max}$  then  $\mathcal{E} \notin \succeq_{max}$ .

*Proof Property 9.* Let us suppose that there exists a transitive relation which satisfies  $P1$  and  $P5$  and which generalizes stable semantics. Let us now consider a system depicted in Figure 2. Suppose that attacks are as depicted and that  $\geq = \{(w, w) \mid w \in \mathcal{A}\}$ . From  $P1$ , we have that for any  $\mathcal{E}' \notin \mathcal{CF}$ , it holds that  $\{x\} \succeq \mathcal{E}'$ . From  $P5$ ,  $\{x\} \succeq$



**Fig. 2** No transitive relation generalizes stable semantics and verifies  $P1$  and  $P5$ .

$\{a\}, \{a\} \succeq \{x\}, \{x\} \succeq \{y\}, \{y\} \succeq \{z\}, \{x\} \succeq \emptyset$ . From those relations and transitivity of  $\succeq$ , we have  $\{x\} \succeq \{x\}$  and  $\{x\} \succeq \{z\}$ . Thus,  $\{x\} \in \succeq_{max}$ . This contradicts the fact that  $\succeq$  generalizes stable semantics, since  $\{x\}$  is not a stable extension of the system  $(\mathcal{A}, \mathcal{R})$ .

*Proof Theorem 8.* We will show that extensions of  $(\mathcal{A}, \mathcal{R})$  coincide with maximal elements of  $\succeq$  for any preference-based argumentation system  $\mathcal{T}$ , such that  $(\nexists a, b \in \mathcal{A})(a, b) \in \mathcal{R} \wedge (b, a) \in \triangleright$ .

$\Rightarrow$  Let  $\mathcal{E} \in \text{Ext}(\mathcal{F})$ . We prove that  $\mathcal{E} \in \succeq_{max}$ . Let  $\mathcal{E}' \in \mathcal{P}(\mathcal{A})$ . If  $\mathcal{E}' \notin \mathcal{CF}$  then, from Postulate 1,  $\mathcal{E} \succeq \mathcal{E}'$ . Let  $\mathcal{E}' \in \mathcal{CF}$ . Since  $\mathcal{E} \in \text{Ext}(\mathcal{F})$  then  $(\forall x' \in \mathcal{E}' \setminus \mathcal{E})(\exists x \in \mathcal{E} \setminus \mathcal{E}')(x, x') \in \mathcal{R}$ . We supposed  $(\nexists a, b \in \mathcal{A})(a, b) \in \mathcal{R} \wedge (b, a) \in \succ$ . Thus, from Postulate 5,  $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$ . Now, Postulate 6 implies  $\mathcal{E} \succeq \mathcal{E}'$ . Since  $\mathcal{E}'$  was arbitrary, then  $\mathcal{E} \in \succeq_{max}$ .

$\Leftarrow$  Let  $\mathcal{E} \in \succeq_{max}$ . We will show that  $\mathcal{E} \in \text{Ext}(\mathcal{F})$ . From Postulate 1,  $\mathcal{E} \in \mathcal{CF}$ . Let  $x' \notin \mathcal{E}$ . Since  $\mathcal{E} \in \succeq_{max}$  then it must be  $\mathcal{E} \succeq \{x'\}$ . From Postulate 4,  $(\exists x \in \mathcal{E})(x, x') \in \mathcal{R} \vee (x, x') \in \succ$ . 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 4,  $\neg(\mathcal{E} \setminus X \succeq \{x'\})$ . This fact and Postulate 6 imply  $\neg(\mathcal{E} \succeq (X \cup \{x'\}))$ . Contradiction with  $\mathcal{E} \in \succeq_{max}$ . Thus,  $\mathcal{E} \in \text{Ext}(\mathcal{F})$ .

*Proof Property 10.* It is easy to show that relation  $\succeq_{gn}$  satisfies  $P1, P4, P5$  and  $P6$ . Postulate 1 is satisfied since from the first item of the definition of  $\succeq_{gn}$ , any conflict-free set is preferred to any conflicting set. Postulate 6 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. Postulate 4 implies that the second item of Definition 16 is not satisfied. Postulate 5 is trivially verified.

*Proof Property 11.* Let us show that  $\succeq_{sp}$  satisfies  $P1, P4, P5$  and  $P6$ . We see from the first item of Definition 17 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 6, 4 and 5 are verified for the same reasons as in the case of relation  $\succeq_{gn}$ .

*Proof Theorem 9.* We will show that for any relation  $\succeq$  which satisfies  $P1, P4, P5$  and  $P6$ , we have that if  $\mathcal{E} \succeq_{gn} \mathcal{E}'$  then  $\mathcal{E} \succeq \mathcal{E}'$  and if  $\mathcal{E} \succeq \mathcal{E}'$  then  $\mathcal{E} \succeq_{sp} \mathcal{E}'$ .

- Let  $\mathcal{E} \succeq_{gn} \mathcal{E}'$ . 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 6, we have  $\mathcal{E} \succeq \mathcal{E}'$  iff  $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$ . From Definition 16 and Postulate 5,  $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$ . Thus,  $\mathcal{E} \succeq \mathcal{E}'$ .
- If  $\mathcal{E}, \mathcal{E}' \notin \mathcal{CF}(\mathcal{T})$  then, Definition 17 implies  $\mathcal{E} \succeq_{sp} \mathcal{E}'$ . 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 17,  $\mathcal{E} \succeq_{sp} \mathcal{E}'$ . In the non-trivial case, when  $\mathcal{E}, \mathcal{E}' \in \mathcal{CF}(\mathcal{T})$ , from Postulate 6,  $\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E}$ . Suppose that  $\neg(\mathcal{E} \setminus \mathcal{E}' \succeq_{sp} \mathcal{E}' \setminus \mathcal{E})$ . Now, Definition 17 implies  $(\exists x' \in \mathcal{E}' \setminus \mathcal{E})(\nexists x \in \mathcal{E} \setminus \mathcal{E}') \text{ s.t. } ((x, x') \in \succ) \vee ((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ)$ . From this fact and Postulate 4, it holds that  $\neg(\mathcal{E} \setminus \mathcal{E}' \succeq \mathcal{E}' \setminus \mathcal{E})$ . Contradiction.

*Proof Theorem 10.*

We will now prove that a set is a pref-stable extension iff it is conflict-free and its arguments win in all conflicts with exterior ones. Throughout the proof, we will use notation  $\succeq_{max}^{gn}$  to refer to the set of maximal elements wrt. relation  $\succeq_{gn}$ .

Since both relations  $\succeq$  and  $\succeq_{gn}$  verify Postulates 1, 6, 4 and 5, then from Theorem 7,  $\succeq_{max} = \succeq_{max}^{gn}$ . This means that it is sufficient to prove that  $\mathcal{E} \in \succeq_{max}^{gn}$  iff the two conditions of theorem are satisfied.

$\Rightarrow$  Let  $\mathcal{E} \in \succeq_{max}^{gn}$ . Since  $\mathcal{E}$  is a pref-extension, according to Property 1,  $\mathcal{E} \in \mathcal{CF}$ . Let  $x' \in \mathcal{A} \setminus \mathcal{E}$ . We supposed that  $(\nexists a \in \mathcal{A}) \text{ s.t. } (a, a) \in \mathcal{R}$ , so it must be that  $\{x'\}$  is conflict-free. Since  $\mathcal{E} \in \succeq_{max}^{gn}$ , it holds that  $\mathcal{E} \succeq_{gn} \{x'\}$ . Since  $\mathcal{E}$  and  $\{x'\}$  are conflict-free, Definition 16 implies  $(\exists x \in \mathcal{E}) \text{ s.t. } (((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \succ))$ .

$\Leftarrow$  Let  $\mathcal{E}$  be conflict-free set and let  $(\forall x' \in \mathcal{A} \setminus \mathcal{E}) (\exists x \in \mathcal{E}) \text{ s.t. } (((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \succ))$ . Let us prove that  $\mathcal{E} \in \succeq_{max}^{gn}$ .

- Since  $\mathcal{E} \in \mathcal{CF}$  then for every non conflict-free set  $\mathcal{E}'$  it holds that  $\mathcal{E} \succeq_{gn} \mathcal{E}'$ .
- 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  $(\exists x \in \mathcal{E} \setminus \mathcal{E}') \text{ s.t. } ((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ)$  or  $((x', x) \in \mathcal{R} \wedge (x, x') \in \succ)$ . Thus,  $\mathcal{E} \succeq_{gn} \mathcal{E}'$ .

From those two items, we have that  $\mathcal{E} \in \succeq_{max}^{gn}$ .

*Proof Theorem 11.*

We will show that a set is a pref-stable extension iff it is conflict-free and attacks wrt.  $\mathcal{R}'$  all arguments in its exterior. Throughout the proof, we will use notation  $\succeq_{max}^{gn}$  to refer to the set of maximal elements wrt relation  $\succeq_{gn}$ .

Since both relations  $\succeq$  and  $\succeq_{gn}$  verify Postulates 1, 6, 4 and 5, then from Theorem 7,  $\succeq_{max} = \succeq_{max}^{gn}$ . This means that it is sufficient to prove that  $\mathcal{E} \in \succeq_{max}^{gn}$  iff  $\mathcal{E} \in \text{Ext}(\mathcal{A}, \mathcal{R}')$ . 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  $\mathcal{E} \in \succeq_{max}^{gn}$ . From Theorem 10,  $\mathcal{E} \in \mathcal{CF}$  and  $(\forall x' \in \mathcal{A} \setminus \mathcal{E}) (\exists x \in \mathcal{E}) \text{ s.t. } (((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \succ))$ . This means that  $(\forall x' \in \mathcal{A} \setminus \mathcal{E}) (\exists x \in \mathcal{E}) \text{ s.t. } (x, x') \in \mathcal{R}'$ . In other words,  $\mathcal{E} \in \text{Ext}(\mathcal{A}, \mathcal{R}')$ .

$\Leftarrow$  Let  $\mathcal{E} \in \text{Ext}(\mathcal{A}, \mathcal{R}')$ . 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  $\mathcal{E} \in \text{Ext}(\mathcal{A}, \mathcal{R}')$ , then  $(\forall x' \in \mathcal{A} \setminus \mathcal{E}) (\exists x \in \mathcal{E}) (x, x') \in \mathcal{R}'$ . This is equivalent to  $(\forall x' \in \mathcal{A} \setminus \mathcal{E}) (\exists x \in \mathcal{E}) \text{ s.t. } (((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \succ))$ . Trivially,  $(\forall x' \in \mathcal{E}' \setminus \mathcal{E}) (\exists x \in \mathcal{E} \setminus \mathcal{E}') \text{ s.t. } (((x, x') \in \mathcal{R} \wedge (x', x) \notin \succ) \vee ((x', x) \in \mathcal{R} \wedge (x, x') \in \succ))$ . That means that  $\mathcal{E} \in \succeq_{max}$ .

*Proof Theorem 12.* Let  $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \succeq)$  be a PAF and  $\mathcal{B} \subseteq \mathcal{A}$ . Let  $x\mathcal{R}'y$  be defined as  $(x\mathcal{R}y$  and not  $(y > x)$ ) or  $(y\mathcal{R}x$  and  $x > y)$ . Let  $\mathcal{A} = \{a_1, \dots, a_n\}$ . The graph  $(\mathcal{A}, \mathcal{R}')$  can be constructed in polynomial time: it is sufficient to check for any pair  $(a_i, a_j)$  of arguments whether  $(a_i\mathcal{R}a_j$  and not  $(a_j > a_i))$  or  $(a_j\mathcal{R}a_i$  and  $a_i > a_j)$  holds or not. In Theorem 11, it has been proved that  $\mathcal{E}$  is a pref-stable extension of  $(\mathcal{A}, \mathcal{R}, \succeq)$  if and only if  $\mathcal{E}$  is a stable extension of  $(\mathcal{A}, \mathcal{R}')$ . Thus, this theorem is a direct consequence of this polynomial transformation and results proved in [31].

*Proof Property 12.*

- We show that  $x \in \mathcal{S}$  iff  $x \in \text{Base}(\text{Arg}(\mathcal{S}))$  where  $\mathcal{S}$  is a consistent subbase of  $\Sigma$ .
  - $\Rightarrow$  Let  $x \in \mathcal{S}$ . Since  $\mathcal{S}$  is consistent, then the set  $\{x\}$  is consistent as well. Thus,  $(\{x\}, x) \in \text{Arg}(\mathcal{S})$ . Consequently,  $x \in \text{Base}(\text{Arg}(\mathcal{S}))$ .
  - $\Leftarrow$  Assume that  $x \in \text{Base}(\text{Arg}(\mathcal{S}))$ . Thus,  $\exists a \in \text{Arg}(\mathcal{S}) \text{ s.t. } x \in \text{Supp}(a)$ . From the definition of an argument,  $\text{Supp}(a) \subseteq \mathcal{S}$ . Consequently,  $x \in \mathcal{S}$ .
- If  $a \in \mathcal{E}$  where  $\mathcal{E} \subseteq \text{Arg}(\Sigma)$ , then  $\text{Supp}(a) \subseteq \text{Base}(\mathcal{E})$ . Consequently,  $a \in \text{Arg}(\text{Base}(\mathcal{E}))$ .
- Let  $\mathcal{S} \subseteq \Sigma$ .
  - Assume that  $\mathcal{S}$  is consistent and  $\text{Arg}(\mathcal{S})$  is not conflict-free. This means that there exist  $a, a' \in \text{Arg}(\mathcal{S}) \text{ s.t. } a\mathcal{R}_{as}a'$ . From Definition 3 of assumption attack, it follows that  $\text{Supp}(a) \cup \text{Supp}(a')$  is inconsistent. Besides, from the definition of an argument,  $\text{Supp}(a) \subseteq \mathcal{S}$  and  $\text{Supp}(a') \subseteq \mathcal{S}$ . Thus,  $\text{Supp}(a) \cup \text{Supp}(a') \subseteq \mathcal{S}$ . Then,  $\mathcal{S}$  is inconsistent. Contradiction.
  - Assume now that  $\mathcal{S}$  is inconsistent. This means that there exists a finite set  $\mathcal{S}' = \{h_1, \dots, h_k\} \text{ s.t.}$ 
    - $\mathcal{S}' \subseteq \mathcal{S}$
    - $\mathcal{S}' \vdash \perp$

- $\mathcal{S}'$  is minimal (wrt. set inclusion) s.t. previous two items hold.

Since  $\mathcal{S}'$  is a minimal inconsistent set, then  $\{h_1, \dots, h_{k-1}\}$  and  $\{h_k\}$  are consistent. Thus,  $(\{h_1, \dots, h_{k-1}\}, \neg h_k), (\{h_k\}, h_k) \in \mathbf{Arg}(\mathcal{S})$ . Furthermore, those two arguments are conflicting (the former attacks the latter). This means that  $\mathbf{Arg}(\mathcal{S})$  is not conflict-free.

*Proof Theorem 13.* Let  $\mathcal{S}$  be a preferred sub-theory of a knowledge base  $\Sigma$ . Thus,  $\mathcal{S}$  is consistent. From Property 12, it follows that  $\mathbf{Arg}(\mathcal{S})$  is conflict-free. Assume that  $\exists a \notin \mathbf{Arg}(\mathcal{S})$ . Since  $a \notin \mathbf{Arg}(\mathcal{S})$  and  $\mathcal{S}$  is a maximal consistent subbase of  $\Sigma$  (according to [21]),  $\exists h \in \mathbf{Supp}(a)$  s.t.  $\mathcal{S} \cup \{h\} \vdash \perp$ . Assume that  $h \in \Sigma_j$ . Thus,  $\mathbf{Level}(\mathbf{Supp}(a)) \geq j$ .

Since  $\mathcal{S}$  is a preferred sub-theory of  $\Sigma$ , then  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_j$  is a maximal (for set inclusion) consistent subbase of  $\Sigma_1 \cup \dots \cup \Sigma_j$ . Thus,  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_j \cup \{h\} \vdash \perp$ . This means that there exists an argument  $(\mathcal{S}', \neg h) \in \mathbf{Arg}(\mathcal{S})$  s.t.  $\mathcal{S}' \subseteq \mathcal{S}_1 \cup \dots \cup \mathcal{S}_j$ . Thus,  $\mathbf{Level}(\mathcal{S}') \leq j$ . Consequently,  $(\mathcal{S}', \neg h) \geq_{WLP} a$ . Moreover,  $(\mathcal{S}', \neg h) \mathcal{R}_{as} a$ . Thus,  $(\mathcal{S}', \neg h) \mathcal{R}_{as} a$ . The second part of the theorem follows directly from Property 12.

*Proof Theorem 14.* Throughout the proof, we will use the notation  $\mathcal{S}_i = \mathcal{S} \cap \Sigma_i$ .

- We will first show that if  $\mathcal{S} \subseteq \Sigma$ ,  $\mathcal{E} = \mathbf{Arg}(\mathcal{S})$  and  $\mathcal{E}$  is a stable extension then  $\mathcal{S} \in \mathbf{PST}$ . We will suppose that  $\mathcal{S}$  is not a preferred sub-theory and we will prove that  $\mathcal{E}$  is not a stable extension. If  $\mathcal{S}$  is not consistent, then Property 12 implies that  $\mathcal{E}$  is not conflict-free. Let us study the case when  $\mathcal{S}$  is consistent but it is not a preferred subtheory. Thus, there exists  $i \in \{1, \dots, n\}$  such that  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_i$  is not a maximal consistent set in  $\Sigma_1, \dots, \Sigma_i$ . Let  $i$  be minimal s.t.  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_i$  is not a maximal consistent set in  $\Sigma_1, \dots, \Sigma_i$ . This means that there exists  $x \notin \mathcal{S}$  s.t.  $x \in \Sigma_i$  and  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_i \cup \{x\}$  is consistent. Let  $a' = (\{x\}, x)$ . Since  $\mathcal{E}$  is a stable extension, then  $(\exists a \in \mathcal{E})$  s.t.  $a \mathcal{R}_r a'$ . Since  $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_i \cup \{x\}$  is consistent then no argument in  $\mathcal{E}$  having level at most  $i$  cannot be in conflict with  $a'$ . Thus, we have that  $\nexists a \in \mathcal{E}$  s.t.  $a \mathcal{R}_r a'$ , which proves that  $\mathcal{E}$  is not a stable extension.
- We will now prove that if  $\mathcal{E} \subseteq \mathcal{A}$  is a stable extension of  $(\mathcal{A}, \mathcal{R}, \geq)$  and  $\mathcal{S} = \mathbf{Base}(\mathcal{E})$  then  $\mathcal{E} = \mathbf{Arg}(\mathcal{S})$ . Suppose the contrary. From Property 12,  $\mathcal{E} \subseteq \mathbf{Arg}(\mathbf{Base}(\mathcal{E}))$ , thus  $\mathcal{E} \subsetneq \mathbf{Arg}(\mathbf{Base}(\mathcal{E}))$ .
  - Let us suppose that  $\mathcal{S}$  is consistent. Since  $\mathcal{S}$  is consistent, then Property 12 implies that  $\mathbf{Arg}(\mathcal{S})$  is conflict-free. Since we supposed that  $\mathcal{E} \subsetneq \mathbf{Arg}(\mathcal{S})$ , then  $\mathcal{E}$  is not maximal conflict-free, contradiction.
  - Let us study the case when  $\mathcal{S}$  is inconsistent. This means that there can be found a set  $\mathcal{S}' = \{h'_1, \dots, h'_k\}$  s.t.
    - $\mathcal{S}' \subseteq \mathcal{S}$
    - $\mathcal{S}' \vdash \perp$
    - $\mathcal{S}'$  is a minimal s.t. the previous two conditions are satisfied.

Let us consider the set  $\mathcal{E}'$  containing the following  $k$  arguments:  $\mathcal{E}' = \{a'_1, \dots, a'_k\}$ , where  $a'_i = (\mathcal{S}' \setminus h'_i, \neg h'_i)$ . Since  $(\forall h'_i \in \mathcal{S}') (\exists a \in \mathcal{E})$  s.t.  $h'_i \in \mathbf{Supp}(a)$  and since  $\mathcal{E}$  is conflict-free then  $(\nexists b \in \mathcal{E})$  s.t.  $\mathbf{Conc}(b) \in \{\neg h'_1, \dots, \neg h'_k\}$ . Hence,  $(\forall a'_i \in \mathcal{E}')$  we have that  $a'_i \notin \mathcal{E}$ . Formally,  $\mathcal{E} \cap \mathcal{E}' = \emptyset$ . This also means that, wrt.  $\mathcal{R}$ , no argument in  $\mathcal{E}$  attacks any of arguments  $a'_1, \dots, a'_k$ . Formally,  $(\forall a' \in \mathcal{E}') (\nexists a \in \mathcal{E})$  s.t.  $a \mathcal{R} a'$ . Since  $\mathcal{E}$  is a stable extension then arguments of  $\mathcal{E}'$  must be attacked wrt.  $\mathcal{R}_r$ . We have just seen that they are not attacked wrt.  $\mathcal{R}$ . This means that:

$$(\forall i \in \{1, \dots, k\}) (\exists a_i \in \mathcal{E}) (a'_i \mathcal{R} a_i) \wedge (a_i > a'_i).$$

For assumption attacks to exist, it is necessary that:

$$(\forall i \in \{1, \dots, k\}) (h'_i \in \text{Supp}(a_i)) \wedge (a_i > a'_i).$$

From  $(\forall i \in \{1, \dots, k\}) a_i > a'_i$  we have  $(\forall i \in \{1, \dots, k\}) \text{Level}(\{h_i\}) \leq \text{Level}(\text{Supp}(a_i)) < \text{Level}(\text{Supp}(a'_i))$ . This means that:

$$(\forall i \in \{1, \dots, k\}) \text{Level}(\{h'_i\}) < \max_{j \neq i} \text{Level}(\{h'_j\}).$$

Let  $l_i = \text{Level}(h'_i)$ , for all  $i \in \{1, \dots, k\}$  and let  $l_m \in \mathcal{S}'$  be s.t.  $l_m = \max\{l_1, \dots, l_k\}$ . Then, from the previous facts, we have:

$$\begin{aligned} l_1 &< l_m \\ &\dots \\ l_m &< \max(\{l_1, \dots, l_k\} \setminus \{l_m\}) \\ &\dots \\ l_k &< l_m \end{aligned}$$

The row  $m$ , i.e.  $l_m < \max(\{l_1, \dots, l_k\} \setminus \{l_m\})$  is an obvious contradiction since we supposed that  $l_m$  is the maximal value in  $\{l_1, \dots, l_k\}$ .

– Now, we have proved that:

1. If  $\mathcal{S} \subseteq \Sigma$ ,  $\mathcal{E} = \text{Arg}(\mathcal{S})$  and  $\mathcal{E}$  is a stable extension, then  $\mathcal{S}$  is a preferred sub-theory.
2. If  $\mathcal{E}$  is a stable extension then  $\mathcal{E} = \text{Arg}(\text{Base}(\mathcal{E}))$ .

Let  $\mathcal{E}$  be a stable extension and let  $\mathcal{S} = \text{Base}(\mathcal{E})$ . Then, from (2),  $\mathcal{E} = \text{Arg}(\mathcal{S})$ . From (1),  $\mathcal{S}$  is a preferred sub-theory.

## References

1. L. Amgoud and Ph. 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 Ph. Besnard. A formal analysis of logic-based argumentation systems. In *International Conference on Scalable Uncertainty Management (SUM'10)*, pages 42–55, 2010.
3. L. Amgoud, Ph. Besnard, and S. Vesic. Identifying the core of logic-based argumentation systems. In *Proceedings of the International Conference on Tools with Artificial Intelligence (ICTAI'11)*, 2011.
4. L. Amgoud, M. Caminada, C. Cayrol, , MC. Lagasquie, and H. Prakken. Towards a consensual formal model: inference part. Technical report, In Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making. ASPIC project, 2004.
5. L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, 29 (2):125–169, 2002.
6. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.
7. L. Amgoud, C. Cayrol, and D. LeBerre. Comparing arguments using preference orderings for argument-based reasoning. In *Proceedings of the 8th International Conference on Tools with Artificial Intelligence (ICTAI'96)*, pages 400–403, 1996.
8. L. Amgoud, Y. Dimopoulos, and P. Moraitis. Making decisions through preference-based argumentation. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 113–123, 2008.
9. L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS'00)*, pages 31–38, 2000.

10. L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence Journal*, 173:413–436, 2009.
11. L. Amgoud and M. Serrurier. Agents that argue and explain classifications. *Journal of Autonomous Agents and Multi-Agents Systems*, 16:187–209, 2008.
12. L. Amgoud and S. Vesic. Repairing preference-based argumentation systems. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'09)*, pages 665–670, 2009.
13. L. Amgoud and S. Vesic. Generalizing stable semantics by preferences. In *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA'10)*, pages 39–50, 2010.
14. L. Amgoud and S. Vesic. On the role of preferences in argumentation frameworks. In *Proceedings of the 22nd International Conference on Tools with Artificial Intelligence (ICTAI'10)*, pages 219–222, 2010.
15. P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence Journal*, 171:675–700, 2007.
16. P. Baroni, M. Giacomin, and G. Guida. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence Journal*, 168:162–210, 2005.
17. T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
18. S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 411–419, 1993.
19. Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.
20. B. Bonet and H. Geffner. Arguing for decisions: A qualitative model of decision making. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI'96)*, pages 98–105, 1996.
21. G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'89)*, pages 1043–1048, 1989.
22. G. Brewka and T. Eiter. Preferred answer sets for extended logic programs. *Artificial Intelligence Journal*, 109:297–356, 1999.
23. G. Brewka, I. Niemela, and M. Truszczynski. Preferences and nonmonotonic reasoning. *AI Magazine*, pages 69–78, 2008.
24. M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence Journal*, 171 (5-6):286–310, 2007.
25. M.W.A. Caminada. Semi-stable semantics. In *Proceedings of the 1st International Conference on Computational Models of Argument (COMMA'06)*, pages 121–130, 2006.
26. C. Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, pages 1443–1448, 1995.
27. C. Cayrol, V. Royer, and C. Saurel. Management of preferences in assumption-based reasoning. *Lecture Notes in Computer Science*, 682:13–22, 1993.
28. Y. Dimopoulos, P. Moraitis, and L. Amgoud. Extending argumentation to make good decisions. In *Proceedings of the 1st International Conference on Algorithmic Decision Theory (ADT'09) – LNCS 5783*, pages 225–236, 2009.
29. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and  $n$ -person games. *Artificial Intelligence Journal*, 77:321–357, 1995.
30. P.M. Dung, P. Mancarella, and F. Toni. Computing ideal skeptical argumentation. *Artificial Intelligence Journal*, 171:642–674, 2007.
31. P. Dunne. Computational properties of argument systems satisfying graph-theoretic constraints. *Artificial Intelligence Journal*, 171 (10-15):701–729, 2007.
32. M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as ‘logical uncertainty’. In *Proceedings of the 2nd European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'93)*, pages 85–90, 1993.
33. S. Kaci. Refined preference-based argumentation frameworks. In *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA'10)*, pages 299–310, 2010.
34. S. Kaci and L. van der Torre. Preference-based argumentation: Arguments supporting multiple values. *Journal of Approximate Reasoning*, 48(3):730751, 2008.

35. S. Kaci, L. van der Torre, and E. Weydert. Acyclic argumentation: Attack = conflict + preference. In *Proceedings of the European Conference on Artificial Intelligence (ECAI'06)*, pages 725–726, 2006.
36. S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. *Journal of Artificial Intelligence*, 104:1–69, 1998.
37. G. Martinez, A. Garcia, and G. Simari. On defense strength of blocking defeaters in admissible sets. In *Proceedings of KSEM'07*, pages 140–152. LNAI 4798, 2007.
38. S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence Journal*, 173:9–10:901–934, 2009.
39. Sanjay Modgil and Henry Prakken. Revisiting preferences and argumentation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 1021–1026, 2011.
40. M. Mozina, J. Zabkar, and I. Bratko. Argument based machine learning. *Artificial Intelligence Journal*, 171(10-15):922–937, 2007.
41. J. Pollock. How to reason defeasibly. *Artificial Intelligence Journal*, 57:1–42, 1992.
42. H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.
43. H. Prakken. An abstract framework for argumentation with structured arguments. *Journal of Argument and Computation*, page In Press, 2011.
44. H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.
45. N. Rescher and R. Manor. On inference from inconsistent premises. *Journal of Theory and decision*, 1:179–219, 1970.
46. G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence Journal*, 53:125–157, 1992.
47. K. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.
48. A. Tarski. *On Some Fundamental Concepts of Metamathematics*. Logic, Semantics, Metamathematic. Edited and translated by J. H. Woodger, Oxford Uni. Press, 1956.