Rich preference-based argumentation frameworks

Leila Amgoud, Srdjan Vesic

Rich preference-based argumentation frameworks ☆

Leila Amgoud [a,*], Srdjan Vesic [b]

[a] IRIT, CNRS, France
[b] CRIL, CNRS, France

A B S T R A C T

An *argumentation framework* is seen as a directed graph whose nodes are *arguments* and arcs are *attacks* between the arguments. Acceptable sets of arguments, called *extensions*, are computed using a *semantics*. Existing semantics are solely based on the attacks and do not take into account other important criteria like the intrinsic *strengths* of arguments.
The contribution of this paper is three fold. First, we study how *preferences* issued from differences in strengths of arguments can help in argumentation frameworks. We show that they play two distinct and complementary roles: (i) to repair the attack relation between arguments, (ii) to refine the evaluation of arguments. Despite the importance of both roles, only the first one is tackled in existing literature. In a second part of this paper, we start by showing that existing models that repair the attack relation with preferences do not perform well in certain situations and may return counter-intuitive results. We then propose a new abstract and general framework which treats properly both roles of preferences. The third part of this work is devoted to defining a bridge between the argumentation-based and the coherence-based approaches for handling inconsistency in knowledge bases, in particular when priorities between formulae are available. We focus on two well-known models, namely the preferred sub-theories introduced by Brewka and the demo-preferred sets defined by Cayrol, Royer and Saurel. For each of these models, we provide an instantiation of our abstract framework which is in full correspondence with it.

1. Introduction

Argumentation is a reasoning model based on the construction and the evaluation of interacting *arguments*. An argument is seen as a reason for believing in a statement, doing an action, pursuing a goal, etc. Argumentation has gained an increasing interest from researchers in Artificial Intelligence. It has, for instance, been used for handling inconsistency in knowledge bases (e.g. [4–6]), merging several knowledge bases (e.g. [7]), making decisions under uncertainty (e.g. [8–10]), modeling different types of dialogs, namely negotiation (e.g. [11,12]) and persuasion [13–15].

One of the most popular argumentation frameworks was proposed by Dung [16]. It consists of a set of arguments and an *attack* relation among them. The framework can thus be represented as a directed graph whose nodes are the arguments and the arcs represent the attacks. Arguments are evaluated using a *semantics* which is based solely on the attack relation and on a key principle according to which an attacker wins unless the attacked argument is defended by other “good” arguments.

In argumentation literature and since early nineties, in their seminal paper [6], Simari and Loui have emphasized the importance of considering additional criteria, namely *preferences*, when evaluating arguments in a framework. Preferences

☆ This paper extensively develops and extends the contents of three conference papers, namely [1–3].
* Corresponding author.
E-mail addresses: amgoud@irit.fr (L. Amgoud), vesic@cril.fr (S. Vesic).

are expressed between arguments and reflect their relative strengths. They may have different sources, like available priorities between formulas of a knowledge base over which an argumentation framework is built [6,17] or the importance of values that may be promoted by arguments [18], etc. In [6], the authors used a weaker version of the basic principle of Dung's semantics. They argued that attacks do not always win. They proposed to remove from an argumentation graph any *critical attack*, i.e., an arc that emanates from an argument which is weaker, or less preferred, than the argument it attacks. This idea was largely acknowledged in the literature and was later applied by several scholars to Dung's abstract framework [17–20] and to its logic-based instantiations [21,22].

Note that while the previous idea is very intuitive, preferences do not play any role in an argumentation framework which does not have critical attacks. In [23], an interesting example of such a framework was given. The framework has two extensions and one of them is clearly better than the other since its arguments are preferred to those of the second extension. This suggests that handling critical attacks is not the unique role that preferences may play in an argumentation framework. They may be used to refine the set of extensions of a framework. Unfortunately, this second role is completely neglected in the literature and existing frameworks capture only the first role. Things are different for the other approaches for defeasible reasoning (e.g., [24–26]). Indeed, priorities between formulas or defaults are used in a model in order to select *some* solutions among possible ones. Thus, the second role is the most popular in these models.

In this paper, we investigate the different roles that preferences may play in a Dung style argumentation framework. We show that there are indeed two distinct roles: handling critical attacks and refining the results of a framework. The two roles are *independent* in the sense that none of the roles can be captured by the other. Moreover, they are not modeled in the same way. We provide one model for each role. Regarding handling critical attacks, the existing approach which consists of deleting them from an argumentation graph may have serious problems in some cases. Indeed, if the attack relation is not symmetric, removing arcs may lead to conflicting extensions and consequently the framework may violate the consistency postulate proposed in [27]. So, instead of deleting the critical attacks, we propose to invert their arrows keeping thus the information about the conflicts and in the same time considering the preferences. We show that this novel approach is well-founded, i.e., it guarantees safe and intuitive results.

For refining the results of an argumentation framework, namely its set of extensions under a given semantics, we use a preference relation defined on the powerset of the set of arguments. The best extensions with respect to this relation will be kept and the others are discarded. Note that such a relation is not unique and two relations may lead to different outcomes for the same argumentation framework. In the paper, we do not study all possible relations. An interesting contribution on defining preference relations between sets can be found in [28].

Another important contribution of the paper consists of proposing a unified abstract argumentation framework that extends Dung's framework with preferences, and more importantly that captures the two roles of preferences. Note that this is the first framework that captures both roles in the same time. It starts by inverting the arrows of the critical attacks, then computes the extensions of the revised graph, and finally applies a refinement relation on the set of extensions in order to select the best ones.

The last contribution of the paper consists of applying our rich framework for reasoning about inconsistency. Our aim is to make bridges with well-known approaches, namely the preferred sub-theories introduced by Brewka in [29] and the demo-preferred sets defined by Cayrol, Royer and Saurel in [30]. Each of these models is in full correspondence with a particular instantiation of the rich model.

The paper is organized as follows: Section 2 recalls Dung's framework. Section 3 discusses the two roles of preferences and how they should be modeled. Section 4 presents an abstract framework which handles the two roles of preferences. Section 5 presents briefly existing approaches [29,30] for handling inconsistency. In Section 6, we show some links between a knowledge base and the arguments which can be built over it. A bijection between preferred sub-theories and a particular case of the framework we propose is shown. This result is generalized by presenting a link between demo-preferred sets and more general instantiation of our rich preference-based framework. Section 7 compares our approach with existing works. The last section concludes.

2. Basics of argumentation

Dung has developed one of the most abstract argumentation framework in the literature [16]. It consists of a set of arguments and an attack relation between the arguments.

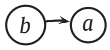
Definition 1 *(Argumentation framework).* An *argumentation framework* (AF) is a pair $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a set of arguments and $\mathcal{R}$ is an attack relation ($\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$). The notation $a\mathcal{R}b$ or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

Let us consider the following example borrowed from [31].

Example 1. Assume the following dialog between an expert and a three-years old child.

Expert: This violin is expensive since it was made by Stradivari (a).
Child: The violin was not made by Stradivari (b).

The corresponding argumentation framework $\mathcal{F}_1$ is depicted in the figure below:



In Dung's framework, arguments and attacks are abstract entities, thus neither their origin nor their structure are known. In the logic-based instantiations of the framework like the one proposed in [21] for reasoning about inconsistent propositional knowledge bases, arguments are built from a knowledge base and are considered as minimal proofs for formulas.

Definition 2 *(Argument).* Let Σ be a finite propositional knowledge base. An *argument* is a pair $\alpha = (H, h)$ such that:

- $H \subseteq \Sigma$.
- H is consistent.
- $H \vdash h$ (where $\vdash$ stands for the classical entailment).
- $\nexists H' \subset H$ such that $H' \vdash h$.

H is the support of the argument and h its conclusion.

The first condition shows the origin of the support of an argument. The second condition ensures the consistency of the support. The third condition says that the conclusion of an argument is a consequence of its support. Finally, the last condition ensures that an argument uses only the necessary information in order to draw its conclusion.

The attack relation is the second key component of Dung's framework. When applied for reasoning about inconsistent information, this relation captures the logical inconsistency of a knowledge base. An example of relation is the so-called undercut and proposed in [32] (see [33] for more attack relations in case of systems built over propositional knowledge bases).

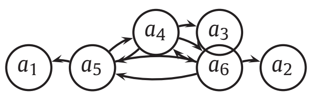
Definition 3 *(Undercut).* An argument (H, h) *undercuts* an argument (H', h') iff $\exists h'' \in H'$ such that $h \equiv \neg h''$.

That is, an argument attacks another one if the conclusion of the first argument contradicts some of the hypothesis of the second one. We illustrate those definitions in the next example.

Example 2. Let $\Sigma = \{x, \neg y, x \to y\}$ be a propositional knowledge base. The following arguments can be built from this base:

a_1: $(\{x\}, x)$	a_2 : $(\{\neg y\}, \neg y)$
a_3: $(\{x \to y\}, x \to y)$	a_4: $(\{x, \neg y\}, x \wedge \neg y)$
a_5: $(\{\neg y, x \to y\}, \neg x)$	a_6: $(\{x, x \to y\}, y)$

Note that more arguments can be built from Σ. The figure below depicts the argumentation framework $\mathcal{F}_2 = (\{a_1, a_2, a_3, a_4, a_5, a_6\}, \mathcal{R})$ where $\mathcal{R}$ is "undercut" between those arguments:



Different *acceptability semantics* for evaluating arguments have been proposed in [16]. Each semantics amounts to define sets of acceptable arguments, called *extensions*. Before recalling those semantics, let us first introduce the two basic properties underlying them, namely *conflict-freeness* and *defence*.

Definition 4 *(Conflict-free, defence).* Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and $\mathcal{E} \subseteq \mathcal{A}$.

- $\mathcal{E}$ is *conflict-free* iff $\nexists a, b \in \mathcal{B}$ such that $a\mathcal{R}b$.
- $\mathcal{E}$ *defends* an argument a iff for all $b \in \mathcal{A}$ such that $b\mathcal{R}a$, there exists $c \in \mathcal{B}$ such that $c\mathcal{R}b$.

The following definition recalls the semantics proposed in [16]. Note that other semantics refining them have been proposed in the literature. However, we do not need to recall them for the purpose of our paper since our approach remains valid for any semantics.

Definition 5 *(Semantics).* Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and $\mathcal{E} \subseteq \mathcal{A}$ is conflict-free.

- $\mathcal{E}$ is an *admissible* extension iff it defends all its elements.
- $\mathcal{E}$ is a *complete* extension iff it is admissible and contains all the arguments it defends.

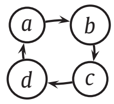
- $\mathcal{E}$ is a *grounded* extension iff it is the minimal (for set inclusion) complete extension.
- $\mathcal{E}$ is a *preferred* extension iff it is a maximal (for set inclusion) admissible extension.
- $\mathcal{E}$ is a *stable* extension iff it attacks any element is $\mathcal{A} \setminus \mathcal{E}$.

Let $\mathtt{Ext}(\mathcal{F})$ denote the set of extensions of $\mathcal{F}$ under a given semantics.

Example 1 *(Cont).* The grounded extension of $\mathcal{F}_1$ is $\{b\}$. This extension is also its unique preferred and stable extension.

Example 2 *(Cont).* The argumentation framework $\mathcal{F}_2$ of Example 2 has an empty grounded extension. However, it has three stable/preferred extensions: $\mathcal{E}_1 = \{a_1, a_2, a_4\}$, $\mathcal{E}_2 = \{a_2, a_3, a_5\}$ and $\mathcal{E}_3 = \{a_1, a_3, a_6\}$.

Example 3. Let us consider the argumentation framework $\mathcal{F}_3$ depicted in the figure below:



$\mathcal{F}_3$ has two preferred and stable extensions: $\{a, c\}$ and $\{b, d\}$. Its grounded extension is the empty set.

The extensions are used for defining the status of each argument. An argument may be either *skeptically* accepted if it belongs to all the extensions of the framework, or *credulously* accepted if it belongs to at least one extension, or *rejected* if it does not belong to any extension.

Definition 6. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and $\mathtt{Ext}(\mathcal{F})$ its set of extensions (under a given semantics). Let $a \in \mathcal{A}$.

- a is *skeptically accepted* iff $\mathtt{Ext}(\mathcal{F}) \neq \emptyset$ and $\forall \mathcal{E}_i \in \mathtt{Ext}(\mathcal{F})$, $a \in \mathcal{E}_i$.
- a is *credulously accepted* iff $\exists \mathcal{E}_i \in \mathtt{Ext}(\mathcal{F})$ such that $a \in \mathcal{E}_i$.
- a is *rejected* iff $\forall \mathcal{E}_i \in \mathtt{Ext}(\mathcal{F})$, $a \notin \mathcal{E}_i$.

Let $\mathtt{Status}(a, \mathcal{F})$ be a function that returns the status of an argument $a \in \mathcal{A}$ in a framework $\mathcal{F}$.

Example 1 *(Cont).* In $\mathcal{F}_1$, the argument b is skeptically accepted while a is rejected under grounded, stable and preferred semantics.

Example 2 *(Cont).* The six arguments of $\mathcal{F}_2$ are all credulously accepted under stable and preferred semantics, and are all rejected under grounded semantics.

Example 3 *(Cont).* In $\mathcal{F}_3$, the four arguments a, b, c, d are credulously accepted under stable and preferred semantics, whereas they are all rejected under grounded semantics.

3. The roles of preferences in argumentation

In this section we investigate the different roles that preferences between arguments may play in an argumentation framework. For each identified role, we propose a framework that handles it properly. Although some properties in this section may look trivial, which is sometimes actually the case as their proofs follow very simple ideas, we state them explicitly in order to show that the system we propose is well-founded. It is also worth mentioning that some of these simple properties are not satisfied by some existing preference-based argumentation frameworks.

In what follows, we assume that $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ is an arbitrary argumentation framework where $\mathcal{A}$ is *finite*. Let $\geqslant$ be a binary relation that expresses preferences between arguments of $\mathcal{A}$. Throughout the paper, the relation $\geqslant \subseteq \mathcal{A} \times \mathcal{A}$ is assumed to be a preorder, i.e., *reflexive* and *transitive*. For two arguments a and b, writing $a \geqslant b$ (or $(a, b) \in \geqslant$) means that a is at least as strong as b. The relation $>$ is the strict version of $\geqslant$. Indeed, $a > b$ iff $a \geqslant b$ and not $(b \geqslant a)$. Finally, $a \approx b$ iff $a \geqslant b$ and $b \geqslant a$. Examples of such relations are those based on the certainty level of the formulae of a propositional knowledge base Σ. For two formulae x and y, writing $x \trianglerighteq y$ means that x is at least as certain as y. We also use the notation $x \triangleright y$ iff $x \trianglerighteq y$ and not $y \trianglerighteq x$. If Σ is equipped with a total preorder $\trianglerighteq$, then it is stratified into $\Sigma_1 \cup \cdots \cup \Sigma_n$ such that formulae of Σ_i have the same certainty level and are more certain than formulae in Σ_j where $j > i$. The stratification of Σ enables to define a certainty level of each subset $\mathcal{S}$ of Σ. It is the highest number of stratum met by this subset. Formally:

$$\mathtt{Level}(\mathcal{S}) = \max\{i \mid \mathcal{S} \cap \Sigma_i \neq \emptyset\} \quad \big(\text{with } \mathtt{Level}(\emptyset) = 0\big).$$

The above certainty level is used by Benferhat, Dubois and Prade [34] in order to define a total preorder on the set of arguments that can be built from a knowledge base. The preorder is defined as follows:

Definition 7 *(Weakest link principle).* (See [34].) Let $\Sigma = \Sigma_1 \cup \cdots \cup \Sigma_n$ be a propositional knowledge. An argument (H, h) is preferred to (H', h'), denoted by $(H, h) \geqslant_{WLP} (H', h')$, iff $\mathtt{Level}(H) \leqslant \mathtt{Level}(H')$.

Example 2 *(Cont).* Assume that $\Sigma = \Sigma_1 \cup \Sigma_2$ with $\Sigma_1 = \{x\}$ and $\Sigma_2 = \{x \to y, \neg y\}$. It holds that $\mathtt{Level}(\{x\}) = 1$ while $\mathtt{Level}(\{\neg y\}) = \mathtt{Level}(\{x \to y\}) = \mathtt{Level}(\{x, \neg y\}) = \mathtt{Level}(\{\neg y, x \to y\}) = \mathtt{Level}(\{x, x \to y\}) = 2$. Thus, $a_1 \geqslant_{WLP} a_2, a_3, a_4, a_5, a_6$ while the five other arguments are all equally preferred.

Let us now analyze the role that preferences between arguments can play in an argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$. We distinguish two roles:

1. To handle correctly the critical attacks in the framework.
2. To refine the evaluation of arguments.

Next subsections discuss in detail each of these roles, their links and how they can be modeled.

3.1. Handling critical attacks

It has been pointed out in the literature that preferences play an important role in argumentation. The idea is that strong arguments are protected against attacks coming from weaker arguments. Let us consider the dialog of Example 1 between an expert and a three-years old child. The argument of the child clearly attacks the argument of the expert. However, this attack should not win since the argument of a child is very weak compared to the argument of an expert. Such attacks conflict with the preference relation between the arguments involved in the attacks. They will be called *critical*.

Definition 8 *(Critical attack).* Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and $\geqslant\ \subseteq\ \mathcal{A} \times \mathcal{A}$. An attack $(b, a) \in \mathcal{R}$ is *critical* iff $a > b$.

In existing literature like [6,35,18,20], critical attacks are removed from argumentation graphs and semantics are applied on reduced graphs. For instance, in Example 1, the arrow from the argument b towards the argument a is removed and Dung's semantics are applied to the new graph.

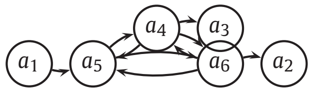
While the idea seems meaningful, removing attacks from an argumentation framework may lead to undesirable situations. Indeed, a reduced argumentation framework may have a conflicting extension. In Example 1, removing the arrow from b to a will lead to a new framework which has a unique stable extension $\{a, b\}$. This extension is not conflict-free and this contradicts the idea that extensions are coherent positions or points of view. In logic-based instantiations of Dung's framework, conflicting extensions will lead to the violation of the rationality postulates discussed in [27,36]. Let us consider the framework $\mathcal{F}_2$ of Example 2. The arrow from a_5 to a_1 is deleted. It can be checked that the set $\{a_1, a_2, a_3, a_5\}$ is a stable extension of the new graph. This extension is clearly not conflict-free. Worse yet, it supports both the formula x (via the argument a_1) and its negation $\neg x$ (via the argument a_5). This means that the framework violates the consistency postulate.

Let us observe that argumentation frameworks that use symmetric attack relations are not concerned by the problem described above since even if an attack is removed, a second attack between the same arguments remains. Thus, the extensions cannot be conflicting. One could say that the problem is solved by using only symmetric attack relations. Unfortunately, this is not possible. Indeed, it was shown in [37] that logic-based frameworks that use a Tarskian logic [38] and symmetric attack relations violate the consistency postulate. This is in particular the case when the knowledge base contains at least one minimal conflict (e.g., $\{x, y, x \to \neg y\}$). Thus, another solution is needed.

The main reason behind the dysfunction of the existing approach is that by removing an attack, a crucial information is lost. In what follows, we propose a novel approach which palliates this limit. The idea is to keep all existing information (arguments and attacks among them). We suggest to modify the graph of attacks by *inverting* the arrow of any critical attack instead of removing it. For instance, in Example 2, the arrow from a_5 to a_1 is replaced by another arrow emanating from a_1 towards a_5. Even if the argument a_1 does not attack the argument a_5 (in the sense of $\mathcal{R}$), it is clear that both arguments cannot be taken together in the same extension. Our approach amounts to taking the strong argument a_1 and discarding a_5. The intuition behind this is that an attack between two arguments represents in some sense two things: (i) an incoherence between the two arguments, and (ii) a kind of preference determined by the direction of the attack. Thus, in our approach, the direction of the arrow represents a real preference between arguments. Moreover, the conflict is kept between the two arguments. Dung's acceptability semantics are then applied on the modified graph.

Definition 9 *(PAF).* A *preference-based argumentation framework* (PAF) is a tuple $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq\ \mathcal{A} \times \mathcal{A}$ is an attack relation and $\geqslant$ is a (partial or total) preorder on $\mathcal{A}$. The extensions of $\mathcal{T}$ under a given semantics are the extensions of the argumentation framework $(\mathcal{A}, \mathcal{R}_r)$, called *repaired framework*, under the same semantics with: $\mathcal{R}_r = \{(a, b) \mid (a, b) \in \mathcal{R} \text{ and not } (b > a)\} \cup \{(b, a) \mid (a, b) \in \mathcal{R} \text{ and } b > a\}$.

Example 2 *(Cont).* In Example 2, we invert the arrow from a_5 to a_1 and obtain the following graph of the repaired framework:

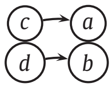


The corresponding PAF has two stable extensions: $\mathcal{E}_1 = \{a_1, a_2, a_4\}$ and $\mathcal{E}_2 = \{a_1, a_3, a_6\}$. We can see that the knowledge base Σ contains three maximal consistent sets: $\{x, x \to y\}$, $\{x, \neg y\}$ and $\{\neg y, x \to y\}$. Since we supposed that x is stronger than the two other formulae, our framework calculates the expected result, namely two stable extensions, both containing an argument for x: one corresponding to the set $\{x, x \to y\}$, and other to $\{x, \neg y\}$. In Section 5, we precisely define a criterion which allows to see why those two maximal consistent sets are better than the third one. In Section 6, we prove that there is a bijection between "the best" maximal consistent sets and the extensions of argumentation framework we propose.

Remark. It is worth mentioning that a full *axiomatics* justification of our solution is provided in [39].

Roughly speaking, one could say that in case of critical attacks, preferences (when they are strict) take precedence over attacks, in the sense that they invert the arrows representing attacks. Let us now check whether the same conclusion holds for non-critical attacks.

Example 4. Let us consider the argumentation framework $\mathcal{F}_4$ depicted in the figure below:



The set $\{c, d\}$ is the only grounded, preferred and stable extension of this argumentation framework. Thus, the two arguments c and d are skeptically accepted while a and b are both rejected.

Assume now that $a > d$ and $b > c$. According to these preferences, it could be argued that the set $\{a, b\}$ is better than $\{c, d\}$. But let us take a closer look at this situation. The argument c attacks a because some fact in a may be challenged by the conclusion of c. Here, the fact that b is stronger than c cannot protect a from c since neither b attacks c nor c attacks b. The simple fact that b is stronger than c cannot protect a. Since a and c are in conflict, one must choose between them. The argument c wins since it attacks a and it is not attacked. Moreover, b and c may be on completely different topics. The same justification holds for choosing d and not b. Thus, one should accept the set $\{c, d\}$.

Informally speaking, we could say that preferences do not take precedence over attacks when they are not critical. Arguments a, b, c, d in the previous example could be given by four witnesses such that c attacks the hypothesis of a and d attacks the hypothesis of b. The fact that witness b is more reliable than c should not be taken into account when reasoning about the conflict between a and c, since argument b is on different topic.

From Definition 9, it is clear that if a PAF has no critical attacks, then the repaired framework coincides with the basic one.

Property 1. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF. If $\mathcal{R}$ has no critical attacks, then $\mathcal{R}_r = \mathcal{R}$. Thus, the extensions of $\mathcal{T}$ are the extensions of $(\mathcal{A}, \mathcal{R})$ under the same semantics.*

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

Example 3 *(Cont).* Consider the framework $\mathcal{F}_3$ and assume that $a > b$ and $c > d$. Note that there are no critical attacks. Thus, the corresponding PAF has two stable extensions: $\{a, c\}$ and $\{b, d\}$.

Example 4 *(Cont).* The attack relation has no critical attacks. Consequently, the stable extension of the PAF is that of $\mathcal{F}_4$, namely $\{c, d\}$.

Our approach does not suffer from the drawback of the existing one. Indeed, it always delivers conflict-free extensions of arguments.

Property 2. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF and $\mathtt{Ext}(\mathcal{T})$ its set of extensions under a given semantics. For all $\mathcal{E}_i \in \mathtt{Ext}(\mathcal{T})$, $\mathcal{E}_i$ is conflict-free wrt. $\mathcal{R}$.*

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

Definition 10. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and $a, b \in \mathcal{A}$. The arguments a and b are *related* in $\mathcal{F}$ iff there exists a finite sequence $a_1, \ldots, a_n$ of arguments such that $n > 1$, $a_1 = a$, $a_n = b$ and for all $i = 1, \ldots, n-1$, either $(a_i, a_{i+1}) \in \mathcal{R}$ or $(a_{i+1}, a_i) \in \mathcal{R}$.

Property 3. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF. For all $a \in \mathcal{A}$ such that $\nexists b, c \in \mathcal{A}$ with $(b, c) \in \mathcal{R}$ is a critical attack and a is related with b, it holds that*:

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

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

Definition 11 *(Maximal elements).* Let $\mathcal{O}$ be a set of objects and $\geqslant \subseteq \mathcal{O} \times \mathcal{O}$ is a preorder. The *maximal elements* of $\mathcal{O}$ wrt. $\geqslant$ are $\texttt{Max}(\mathcal{O}, \geqslant) = \{o \in \mathcal{O} \mid \nexists o' \in \mathcal{O}$ such that $o' > o\}$.

Property 4. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF such that $\geqslant$ is total.*[1]
If $\texttt{Max}(\mathcal{A}, \geqslant)$ is conflict-free (wrt. $\mathcal{R}$), then $\forall a \in \texttt{Max}(\mathcal{A}, \geqslant)$:

- *a is skeptically accepted in $\mathcal{T}$ under preferred and grounded semantics.*
- *if $\mathcal{T}$ has at least one stable extension, then a is skeptically accepted under stable semantics.*

The following result shows that when the preference relation $\geqslant$ is a linear order (i.e. reflexive, antisymmetric, transitive and complete), then the corresponding PAF has a unique stable/preferred extension. Moreover, this extension is computed in $\mathcal{O}(n^2)$ time.

Property 5. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF such that $\mathcal{R}$ is irreflexive and $\geqslant$ is a linear order.*

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

Let us now see what happens in case the attack relation is symmetric. The following result shows that our approach returns the same results as the approach developed in [17,18]. This means that inverting the arrows or removing them will lead to the same result.

Property 6. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF. If $\mathcal{R}$ is symmetric, then* $\texttt{Ext}(\mathcal{T}) = \texttt{Ext}((\mathcal{A}, \mathcal{R}'))$ (*under the same semantics*) *where*

$$\mathcal{R}' = \left\{(a, b) \;\middle|\; (a, b) \in \mathcal{R} \text{ and } \neg(b > a)\right\}.$$

We can also show that when the attack relation is symmetric, the extensions of a PAF are a subset of those of its basic framework. This means that preferences filter the extensions, i.e., help to select only the best ones.

Property 7. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF where $\mathcal{R}$ is symmetric. If $\mathcal{E} \subseteq \mathcal{A}$ is a preferred* (*stable*) *extension of framework $\mathcal{T}$ then $\mathcal{E}$ is a preferred* (*stable*) *extension of $(\mathcal{A}, \mathcal{R})$.*

This property does not hold in case the attack relation is not symmetric as shown in the following example.

[1] A preorder $\geqslant$ on a set $\mathcal{A}$ is total iff for all $a, b \in \mathcal{A}$, $a \geqslant b$ or $b \geqslant a$.

Example 5. The argumentation framework, $\mathcal{F}_5$, depicted in the left side of the figure below has a unique stable extension, $\{b, d\}$. Assume that $a > b$, then the repaired framework is depicted in the right side of the same figure. It can be checked that its stable extension is the set $\{a, c\}$.

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

Property 8. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF such that $\mathcal{R}$ is symmetric, and $\mathcal{E} \subseteq \mathcal{A}$. $\mathcal{E}$ is a stable extension of $(\mathcal{A}, \mathcal{R})$ but not of $\mathcal{T}$ iff $\exists x' \notin \mathcal{E}$ such that $\forall x \in \mathcal{E}$, if $x\mathcal{R}x'$, then $x' > x$.*

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

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

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

3.2. Refining argumentation frameworks by preferences

In the previous subsection, we have studied the case where the attack relation and the preference relation of an argumentation framework are in conflict. We have seen that in case of conflict, the preference relation should take precedence. However, in argumentation frameworks which are free of critical attacks, preferences do not play any role. For instance, the argumentation framework $\mathcal{F}_4$ of Example 4 has no critical attacks and the two preferences $a > d$ ad $b > c$ are completely useless. In this section, we show that preferences play another role in argumentation frameworks. They may be used in order to refine the result of the PAF developed in the previous section. To put it differently, they allow to choose some extensions among the set of extensions of the repaired framework. Let us illustrate our ideas on the following example.

Example 3 *(Cont).* Recall that $\mathcal{F}_3$ has two stable extensions, $\{a, c\}$ and $\{b, d\}$, and the four arguments a, b, c, d are all credulously accepted. Let us now assume that $a > b$ and $c > d$. Note that any element of $\{b, d\}$ is weaker than at least one element of the set $\{a, c\}$. Thus, it is natural to consider $\{a, c\}$ as better than $\{b, d\}$. This can be important in a decision making problem. Assume that the extension $\{c, d\}$ supports an option o_1 while the extension $\{b, d\}$ supports another option, say o_2. Since only one option will be chosen at the end, the available preferences make it possible to select o_1.

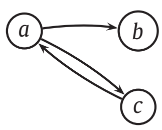
The previous example shows that the extensions of an argumentation framework can be compared on the basis of preferences between arguments. Some extensions may thus be better than others. What is worth noticing is that a refinement amounts to *compare* subsets of arguments. In Example 3, the so-called *democratic* relation, $\succcurlyeq_d$, is used for comparing the two sets $\{a, c\}$ and $\{b, d\}$. This relation is defined as follows:

Definition 12 *(Democratic relation).* Let Δ be a set of objects and $\geqslant \subseteq \Delta \times \Delta$ be a partial preorder. For $\mathcal{X}, \mathcal{X}' \subseteq \Delta$, $\mathcal{X} \succcurlyeq_d \mathcal{X}'$ iff $\forall x' \in \mathcal{X}' \setminus \mathcal{X}$, $\exists x \in \mathcal{X} \setminus \mathcal{X}'$ such that $x > x'$.

There are several other relations which can be used to refine the results of a PAF. For instance, the so-called *elitist* relation is defined as follows.

Definition 13 *(Elitist relation).* Let Δ be a set of objects and $\geqslant \subseteq \Delta \times \Delta$ be a partial preorder. For $\mathcal{X}, \mathcal{X}' \subseteq \Delta$, $\mathcal{X} \succcurlyeq_e \mathcal{X}'$ iff $\forall x \in \mathcal{X} \setminus \mathcal{X}'$, $\exists x' \in \mathcal{X}' \setminus \mathcal{X}$ such that $x > x'$.

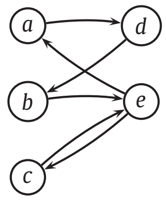
Example 6. Consider the preference-based argumentation framework $\mathcal{T}_6$ with attack relation as depicted in the figure below and suppose that $a > b$:



The corresponding PAF returns two preferred (stable) extensions: $\mathcal{E}_1 = \{a\}$ and $\mathcal{E}_2 = \{b, c\}$. Relation $\succcurlyeq_d$ does not allow to compare those two sets, formally $\neg(\mathcal{E}_1 \succcurlyeq_d \mathcal{E}_2)$ and $\neg(\mathcal{E}_2 \succcurlyeq_d \mathcal{E}_1)$. However, $\mathcal{E}_1 \succ_e \mathcal{E}_2$.

We now provide an example where the converse situation holds.

Example 7. Let us study the PAF $\mathcal{T}_7$ depicted in the figure below with $a > d$ and $b > e$:



The basic PAF returns two preferred (stable) extensions: $\mathcal{E}_1 = \{a, b, c\}$ and $\mathcal{E}_2 = \{d, e\}$. According to democratic relation, $\mathcal{E}_1 \succ_d \mathcal{E}_2$, while $\neg(\mathcal{E}_1 \succcurlyeq_e \mathcal{E}_2)$ and $\neg(\mathcal{E}_2 \succcurlyeq_e \mathcal{E}_1)$.

Note that the first phase, conflict resolution, is responsible for selecting coherent (i.e. conflict-free) points of view which are complete enough in order to attack other sets of arguments, or at least, to defend their own arguments. The second phase, refinement, aims at choosing among those points of view the ones containing the best arguments.

Let us now formally define a refinement relation, i.e., the basic properties that such a relation should satisfy.

Definition 14 *(Refinement relation).* Let $(\mathcal{A}, \geqslant)$ be such that $\mathcal{A}$ is a set of arguments and $\geqslant \subseteq \mathcal{A} \times \mathcal{A}$ is a (partial or total) preorder. A *refinement relation*, denoted by $\succcurlyeq$, is a binary relation on $\mathcal{P}(\mathcal{A})$[2] such that:

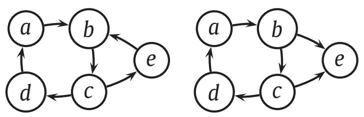
- $\succcurlyeq$ is reflexive.
- $\succcurlyeq$ is transitive.
- For all $\mathcal{E} \subseteq \mathcal{A}$, for all $a, b \in \mathcal{A} \setminus \mathcal{E}$, if $a > b$ then $\mathcal{E} \cup \{a\} \succ \mathcal{E} \cup \{b\}$.

The two first conditions ensure that a refinement relation is a preorder. This is important since a refinement relation plays the role of a preference relation and should thus satisfy some basic properties like reflexivity and transitivity. The third condition ensures a form of monotonicity. It states that if an argument is strictly preferred to another argument, then this preference is preserved by the refinement relation.

Property 10. *Democratic relation and elitist relation are both refinement relations.*

So far, we have shown that preferences play two roles in a PAF: to handle correctly critical attacks and to refine its results. The question now is what are the links between the two roles? Is it possible that one of them subsumes the other? Let us start by studying whether handling correctly critical attacks is sufficient to return "refined" results. The answer is certainly negative since we have shown in a previous section that when a framework has no critical attacks, the available preferences are completely useless. Thus, the result of the framework may still need to be refined. The following example shows that even after repairing $\mathcal{R}$, we still need to refine the results.

Example 8. Let us consider the argumentation framework $\mathcal{F}_8$ depicted in the left side of the following figure:



Assume that $a > b$, $c > d$ and $b > e$. The repaired framework corresponding to $(\mathcal{A}, \mathcal{R}, \geqslant)$ is depicted in the right side of the above figure. This latter has two stable extensions $\{a, c\}$ and $\{b, d\}$. According to the democratic relation $\succcurlyeq_d$, it is clear that the first extension is better than the second one.

It is even more immediate to see that the refinement alone cannot solve the problem of critical attacks. We illustrate this point by the following simple example.

[2] $\mathcal{P}(\mathcal{A})$ denotes the powerset of a set $\mathcal{A}$.

Example 9. Let us consider the argumentation framework $\mathcal{F}_9$ depicted in the left side of the figure below:

If we ignore the critical attacks and preferences, the framework $\mathcal{F}_9$ has two stable/preferred extensions: $\{a, c\}$ and $\{b, d\}$. Assume now that $a > b$ and $a > d$ and $d > c$. When democratic relation $\succcurlyeq_d$ is applied, the set $\{a, c\}$ is preferred to $\{b, d\}$. However, this result is not intuitive. The reason is that d defends itself against its unique attacker c. Thus, d should be accepted and consequently, c should be rejected, thus the expected extension would be: $\{a, d\}$. Note that $\{b, d\}$ cannot be an extension since b is attacked by a stronger argument (a). On the right side of the figure, we see the framework obtained by inverting the arrows of the critical attacks. This framework has a unique stable extension which is the expected result $\{a, d\}$.

The conclusion is that taking in account of preferences is a two-steps process which consists of:

1. Repairing the attack relation $\mathcal{R}$ by computing $\mathcal{R}_r$.
2. Refining the results of the framework $(\mathcal{A}, \mathcal{R}_r)$ by comparing its extensions using a refinement relation.

4. Rich PAFs

In this section we propose an abstract model that extends Dung's argumentation framework with preferences between arguments. The model integrates both roles of preferences. Note that this is the first model that treats the two roles of preferences together. It is also the first model that refines the extensions of argumentation frameworks. The model is referred to as *rich preference-based argumentation framework*.

Definition 15 *(Rich PAF).* A *rich PAF* is a tuple $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant, \succcurlyeq)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation, $\geqslant \subseteq \mathcal{A} \times \mathcal{A}$ is a (partial or total) preorder and $\succcurlyeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ is a refinement relation. The extensions of $\mathcal{T}$ under a given semantics are the elements of $\mathtt{Max}(\mathcal{B}, \succcurlyeq)$ where $\mathcal{B}$ is the set of extensions (under the same semantics) of the PAF $(\mathcal{A}, \mathcal{R}, \geqslant)$.

From the previous definition, it is clear that the extensions of a rich PAF are a subset of the extensions of its repaired framework (thus, a subset of the extensions of the PAF $(\mathcal{A}, \mathcal{R}, \geqslant)$). This means that a rich PAF refines a PAF. Moreover, if the PAF has only one extension, then this latter is the only extension of the rich PAF. Another case where a refinement is not necessary is when the relation $\geqslant$ is a linear order.

Property 11. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant, \succcurlyeq)$ be a rich PAF and $\mathcal{B}$ be the set of extensions (under the same semantics) of the repaired framework $(\mathcal{A}, \mathcal{R}_r)$.*

- $\mathtt{Max}(\mathcal{B}, \succcurlyeq) \subseteq \mathcal{B}$.
- *If $|\mathcal{B}| = 1$, then $\mathtt{Max}(\mathcal{B}, \succcurlyeq) = \mathcal{B}$.*
- *If $\mathcal{R}$ is irreflexive and $\geqslant$ is a linear order, then $\mathtt{Max}(\mathcal{B}, \succcurlyeq) = \mathcal{B}$ holds for stable, preferred, grounded and complete semantics.*

Example 1 *(Cont).* In this example, the PAF has a unique stable extension $\{a\}$. Thus, $\{a\}$ is the unique stable extension of the rich PAF $(\{a, b\}, \{(b, a)\}, \{(a, a), (b, b), (a, b)\}, \succcurlyeq)$ whatever the refinement relation that is used.

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

Property 12. *Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant, \succcurlyeq)$ be a rich PAF such that $\mathcal{R}$ has no critical attacks. Preferred (resp. stable) extensions of $\mathcal{T}$ are exactly the elements of $\mathtt{Max}(\mathcal{S}, \succcurlyeq)$ where $\mathcal{S}$ is the set of all preferred (resp. stable) extensions of the AF $(\mathcal{A}, \mathcal{R})$.*

Example 3 *(Cont).* Let us use the democratic relation $\succcurlyeq_d$. In $\mathcal{F}_3$, there are no critical attacks ($\mathcal{R}_r = \mathcal{R}$). The extensions of the rich PAF are $\mathtt{Max}(\{\{a, c\}, \{b, d\}\}, \succcurlyeq_d) = \{\{a, c\}\}$. Thus, $\{a, c\}$ is the unique stable extension of the corresponding rich PAF.

Example 4 *(Cont).* The PAF extending $\mathcal{F}_4$ has no critical attacks. Moreover, $\mathcal{F}_4$ has a unique stable/preferred extension. Thus, this is also the unique stable/preferred extension of the corresponding rich PAF whatever the refinement relation that is considered.

Example 8 *(Cont).* Recall that the repaired framework of $\mathcal{T}_8$ has two stable extensions: $\{a, c\}$ and $\{b, d\}$. Moreover, $\mathtt{Max}(\{\{a, c\}, \{b, d\}\}, \succcurlyeq_d) = \{\{a, c\}\}$. Thus, $\{a, c\}$ is the unique stable extension of the rich PAF that uses the democratic relation.

5. Coherence-based approach for handling inconsistency

Coherence-based approach for handling inconsistency in a propositional knowledge base Σ follows two steps: At the first step, some subbases of Σ are chosen. These subbases can be, for example, maximal for set inclusion consistent subsets of the knowledge base [40]. At the second step, an inference mechanism is chosen. This later defines the inferences to be made from Σ. 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 Σ is equipped with a (total or partial) preorder $\trianglerighteq$ ($\trianglerighteq \subseteq \Sigma \times \Sigma$). Recall that when $\trianglerighteq$ is total, Σ is stratified into $\Sigma_1 \cup \cdots \cup \Sigma_n$ such that $\forall i, j$ with $i \neq j$, $\Sigma_i \cap \Sigma_j = \emptyset$. Moreover, Σ_1 contains the most important formulas while Σ_n contains the least important ones.

In [29], a knowledge base is equipped with a total preorder and the chosen subbases privilege the most important formulas.

Definition 16 *(Preferred sub-theory).* (See [29].) Let Σ be stratified into $\Sigma_1 \cup \cdots \cup \Sigma_n$. A *preferred sub-theory* is a set $\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_n$ such that $\forall k \in [1, n]$, $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_k$ is a maximal (for set inclusion) consistent subbase of $\Sigma_1 \cup \cdots \cup \Sigma_k$.

Example 2 *(Cont).* The knowledge base $\Sigma = \Sigma_1 \cup \Sigma_2$ with $\Sigma_1 = \{x\}$ and $\Sigma_2 = \{x \to y, \neg y\}$ has two preferred sub-theories: $\mathcal{S}_1 = \{x, x \to y\}$ and $\mathcal{S}_2 = \{x, \neg y\}$.

Brewka [29] has shown that the preferred sub-theories of a knowledge base Σ are maximal (wrt set inclusion) consistent subbases of Σ.

Property 13. *(See [29].) Each preferred sub-theory of a knowledge base Σ is a maximal (for set inclusion) consistent subbase of Σ.*

The above definition has been extended to the case where Σ is equipped with a partial preorder $\trianglerighteq$ [30]. The basic idea was to define a preference relation on the powerset of Σ. The best elements according to this relation are the preferred theories, called also *democratic sub-theories*. The relation that generalizes preferred sub-theories is the democratic relation (see Definition 12). In this context, Δ is Σ and $\geqslant$ is the relation $\trianglerighteq$. In what follows, $\triangleright$ denotes the strict version of $\trianglerighteq$. Thus:

$$\text{Let } \mathcal{S}, \mathcal{S}' \subseteq \Sigma.\ \mathcal{S} \succcurlyeq_d \mathcal{S}' \text{ iff } \forall x' \in \mathcal{S}' \setminus \mathcal{S}, \exists x \in \mathcal{S} \setminus \mathcal{S}' \text{ such that } x \triangleright x'.$$

Definition 17 *(Democratic sub-theory).* (See [30].) Let Σ be propositional knowledge base and $\trianglerighteq \subseteq \Sigma \times \Sigma$ be a partial preorder. A *democratic sub-theory* is a set $\mathcal{S} \subseteq \Sigma$ such that $\mathcal{S}$ is consistent and $\nexists \mathcal{S}' \subseteq \Sigma$ such that $\mathcal{S}'$ is consistent and $\mathcal{S}' \succ_d \mathcal{S}$.

Example 10. Let $\Sigma = \{x, \neg x, y, \neg y\}$ be such that $\neg x \trianglerighteq y$ and $\neg y \trianglerighteq x$. Let $\mathcal{S}_1 = \{x, y\}$, $\mathcal{S}_2 = \{x, \neg y\}$, $\mathcal{S}_3 = \{\neg x, y\}$, and $\mathcal{S}_4 = \{\neg x, \neg y\}$. The three subbases $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ are the democratic sub-theories of Σ. However, $\mathcal{S}_1$ is not a democratic sub-theory since $\mathcal{S}_4 \succ_d \mathcal{S}_1$.

It is easy to show that the democratic sub-theories of a knowledge base Σ are maximal (for set inclusion) consistent.

Property 14. *Each democratic sub-theory of a knowledge base Σ is a maximal (for set inclusion) consistent subbase of Σ.*

6. Computing sub-theories with argumentation

We have already shown in [39] that the first role of preferences (i.e. handling critical attacks) is sufficient to capture the preferred sub-theories of Brewka by a particular argumentation framework. We now show that our rich PAF generalizes this result as it allows to recover the democratic sub-theories. Recall that democratic sub-theories are a generalization of preferred sub-theories. In Subsection 6.1, we recall the results about the link with the preferred sub-theories. Subsection 6.2 presents new results which generalize these results. The two instances (i.e., the two frameworks that recover respectively preferred sub-theories and democratic sub-theories) use all the arguments that can be built from Σ using Definition 2 (i.e. the set $\mathtt{Arg}(\Sigma)$). Similarly, they both use the attack relation "Undercut" given also in Definition 2. However, as we will see next, they are grounded on distinct preference relations between arguments. The last component of a rich PAF is a preference relation on the power set of $\mathtt{Arg}(\Sigma)$. Both instances will use the democratic relation $\succcurlyeq_d$. To sum up, for recovering preferred and democratic sub-theories, we will use two instances of the rich PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant, \succcurlyeq_d)$. Before presenting the formal results, let us first introduce some basic properties.

It can be shown that when the preference relation $\geqslant$ is a total preorder, then the stable extensions of the PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant)$ are all incomparable wrt the democratic relation $\succcurlyeq_d$.

Property 15. *Let $\mathcal{T} = (\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant)$ be a PAF. For all stable extensions $\mathcal{E}$ and $\mathcal{E}'$ of $\mathcal{T}$ with $\mathcal{E} \neq \mathcal{E}'$, if $\geqslant$ is a total preorder, then $\neg(\mathcal{E} \succcurlyeq_d \mathcal{E}')$.*

From the previous property, it follows that the stable extensions of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant)$ coincide with those of the rich PAF $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant, \succcurlyeq_d)$.

Property 16. *If $\geqslant$ is a total preorder, then the stable extensions of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant, \succcurlyeq_d)$ are exactly the stable extensions of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant)$.*

Let us start by introducing some useful notations.

Notations: Let $a = (H, h)$ be an argument (in the sense of Definition 2). The functions $\texttt{Supp}$ and $\texttt{Conc}$ return respectively the support H and the conclusion h of the argument a. For $\mathcal{S} \subseteq \Sigma$, $\texttt{Arg}(\mathcal{S})$ is the set of all arguments that may be built from $\mathcal{S}$ in the sense of Definition 2. For $\mathcal{E} \subseteq \texttt{Arg}(\Sigma)$, $\texttt{Base}(\mathcal{E}) = \bigcup \texttt{Supp}(a)$ where $a \in \mathcal{B}$.

The following result summarizes some useful properties of the above functions.

Property 17.

- *For any consistent subbase $\mathcal{S} \subseteq \Sigma$, $\mathcal{S} = \texttt{Base}(\texttt{Arg}(\mathcal{S}))$.*
- *The function $\texttt{Base}$ is surjective but not injective.*
- *For any $\mathcal{E} \subseteq \texttt{Arg}(\Sigma)$, $\mathcal{E} \subseteq \texttt{Arg}(\texttt{Base}(\mathcal{E}))$.*
- *The function $\texttt{Arg}$ is injective but not surjective.*

Another property that is important for the rest of the paper relates the notion of consistency of a set of formulas to that of conflict-freeness of a set of arguments.

Property 18. *A set $\mathcal{S} \subseteq \Sigma$ is consistent* iff *$\texttt{Arg}(\mathcal{S})$ is conflict-free.*

The following example shows that the previous property does not hold for an arbitrary set of arguments.

Example 11. Let $\mathcal{E} = \{(\{x\}, x), (\{x \to y\}, x \to y), (\{\neg y\}, \neg y)\}$. It is obvious that $\mathcal{E}$ is conflict-free while $\texttt{Base}(\mathcal{E})$ is not consistent.

6.1. Recovering preferred sub-theories

In this subsection, we recall the results we proved in [39] and which show that there is a full correspondence between the preferred sub-theories of a knowledge base Σ and the stable extensions of the PAF $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$. Recall that the relation $\geqslant_{WLP}$ 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. This relation is a total preorder and is defined over a knowledge base that is itself equipped with a total preorder. According to Property 16, the stable extensions of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ coincide with those of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP}, \succcurlyeq_d)$.

The first result shows that from a preferred sub-theory, it is possible to build a unique stable extension of the PAF $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$.

Theorem 1. *(See [39].) Let Σ be a stratified knowledge base. For all preferred sub-theory $\mathcal{S}$ of Σ, it holds that:*

- *$\texttt{Arg}(\mathcal{S})$ is a stable extension of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$.*
- *$\mathcal{S} = \texttt{Base}(\texttt{Arg}(\mathcal{S}))$.*

Similarly, each stable extension of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ is built from a unique preferred sub-theory of Σ.

Theorem 2. *(See [39].) Let Σ be a stratified knowledge base. For all stable extension $\mathcal{E}$ of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$, it holds that:*

- *$\texttt{Base}(\mathcal{E})$ is a preferred sub-theory of Σ.*
- *$\mathcal{E} = \texttt{Arg}(\texttt{Base}(\mathcal{E}))$.*

There exists a one-to-one correspondence between preferred sub-theories of Σ and the stable extensions of $(\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$.

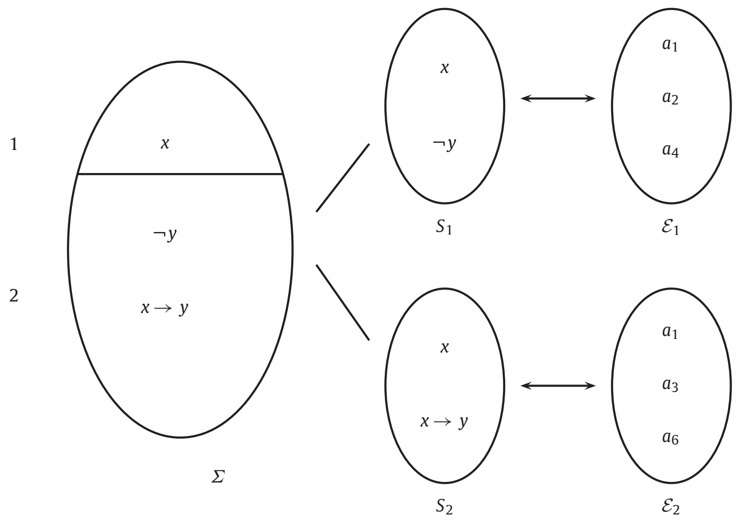


Fig. 1. Preferred sub-theories of Σ vs. stable extensions of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$.

Theorem 3. *(See [39].) Let $\mathcal{T} = (\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ be a PAF over a stratified knowledge base Σ. The stable extensions of $\mathcal{T}$ are exactly the sets $\mathtt{Arg}(\mathcal{S})$ where $\mathcal{S}$ ranges over the preferred sub-theories of Σ.*

From the above result it follows that the PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ has at least one stable extension.

Corollary 1. *If Σ has at least one consistent formula, then the PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ has at least one stable extension.*

Example 2 *(Cont).* Fig. 1 shows the two preferred sub-theories of Σ as well as the two stable extensions of the corresponding PAF.

6.2. Recovering democratic sub-theories

Recall that the democratic sub-theories of a knowledge base Σ generalize the preferred sub-theories when Σ is equipped with a partial preorder $\trianglerighteq$. Thus, in order to capture the democratic sub-theories, we will use the generalized version of the preference relation $\geqslant_{WLP} \subseteq \mathtt{Arg}(\Sigma) \times \mathtt{Arg}(\Sigma)$. The idea behind the new relation, denoted by $\geqslant_{GWLP}$, is that an argument is preferred to another if every formula used in the support of the former is strictly preferred in the sense of $\trianglerighteq$ to at least one formula in the support of the later.

Definition 18 *(Generalized weakest link principle).* (See [21].) Let Σ be a knowledge base which is equipped with a partial preorder $\trianglerighteq$. For two arguments $(H, h), (H', h') \in \mathtt{Arg}(\Sigma)$, $(H, h) \geqslant_{GWLP} (H', h')$ iff $\forall k \in H$, $\exists k' \in H'$ such that $k \triangleright k'$ (i.e. $k \trianglerighteq k'$ and not $(k' \trianglerighteq k)$).

It can be shown that from each democratic sub-theory of a knowledge base Σ, a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$ can be built.

Theorem 4. *Let Σ be a knowledge base which is equipped with a partial preorder $\trianglerighteq$. For all democratic sub-theory $\mathcal{S}$ of Σ, it holds that $\mathtt{Arg}(S)$ is a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$.*

The following result shows that each stable extension of the PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$ returns a maximal consistent subbase of Σ.

Theorem 5. *Let Σ be a knowledge base which is equipped with a partial preorder $\trianglerighteq$. For all stable extension $\mathcal{E}$ of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$, it holds that:*

- *$\mathtt{Base}(\mathcal{E})$ is a maximal (for set inclusion) consistent subbase of Σ.*
- *$\mathcal{E} = \mathtt{Arg}(\mathtt{Base}(\mathcal{E}))$.*

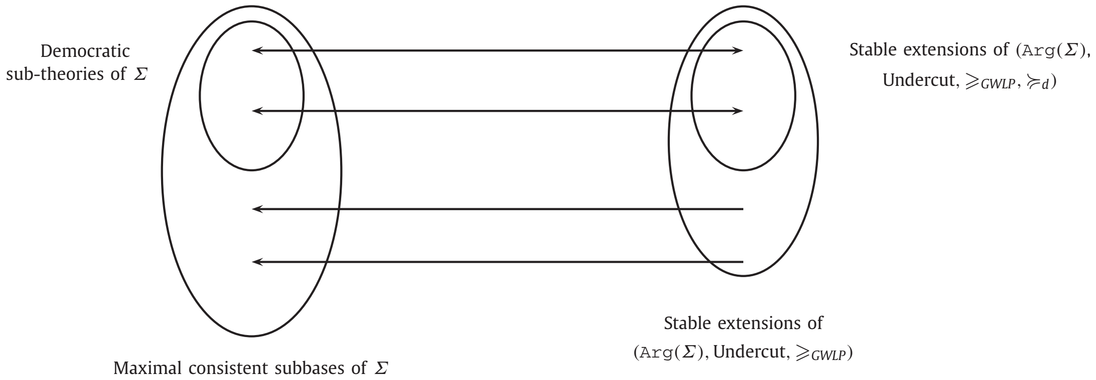


Fig. 2. Democratic sub-theories of Σ vs. Stable extensions of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP}, \succcurlyeq_d)$.

The following example shows that the stable extensions of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$ do not necessarily return democratic sub-theories.

Example 10 *(Cont).* Recall that $\Sigma = \{x, \neg x, y, \neg y\}$, $\neg x \trianglerighteq y$ and $\neg y \trianglerighteq x$. Let $\mathcal{S} = \{x, y\}$. It can be checked that the set $\mathtt{Arg}(\mathcal{S})$ is a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$. However, $\mathcal{S}$ is not a democratic sub-theory since $\{\neg x, \neg y\} \succcurlyeq_d \mathcal{S}$.

It can also be shown that the converse of the above theorem is not true. Indeed, a knowledge base may have a maximal consistent subbase $\mathcal{S}$ such that $\mathtt{Arg}(\mathcal{S})$ is not a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$. Let us consider the following example.

Example 12. Let $\Sigma = \{x, \neg x\}$ and $x \triangleright \neg x$. It is clear that $\{\neg x\}$ is a maximal consistent subbase of Σ while $\mathtt{Arg}(\{\neg x\})$ is not a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$.

The following result establishes a link between the 'best' maximal consistent subbases of Σ wrt the democratic relation $\succcurlyeq_d$ and the 'best' sets of arguments wrt the same relation $\succcurlyeq_d$.

Theorem 6. *Let $\mathcal{S}, \mathcal{S}' \subseteq \Sigma$ be maximal (for set inclusion) consistent subbases of Σ. It holds that $\mathcal{S} \succcurlyeq_d \mathcal{S}'$ iff $\mathtt{Arg}(\mathcal{S}) \succcurlyeq_d \mathtt{Arg}(\mathcal{S}')$.*

We also show that from each democratic sub-theory of Σ, one can build a stable extension of the corresponding rich PAF, and each stable extension of the rich PAF is built from a democratic sub-theory.

Theorem 7. *Let Σ be equipped with a partial preorder $\trianglerighteq$.*

- *For each democratic sub-theory $\mathcal{S}$ of Σ, $\mathtt{Arg}(\mathcal{S})$ is a stable extension of the rich PAF $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP}, \succcurlyeq_d)$.*
- *For each stable extension $\mathcal{E}$ of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP}, \succcurlyeq_d)$, $\mathtt{Base}(\mathcal{E})$ is a democratic sub-theory of Σ.*

Finally, we show that there is a one-to-one correspondence between the democratic sub-theories of a base Σ and the stable extensions of its corresponding rich PAF.

Theorem 8. *The stable extensions of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP}, \succcurlyeq_d)$ are exactly the $\mathtt{Arg}(\mathcal{S})$ where $\mathcal{S}$ ranges over the democratic sub-theories of Σ.*

Fig. 2 synthesizes the different links between the democratic sub-theories of a knowledge base Σ and the stable extensions of its corresponding PAF and rich PAF.

7. Related work

Introducing preferences in argumentation frameworks goes back to Simari and Loui [6]. In that work, the authors have defined an argumentation framework in which arguments are built from a propositional knowledge base. The arguments grounded on specific information are considered as stronger than the ones built from more general information. This preference is used to solve dilemmas between any pair of conflicting arguments. Thus, it is used for handling critical attacks. The idea of this paper has been generalized in [17,18] to any argumentation framework and to any preference relation. Unfortunately, these approaches deliver 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 one is that it may return conflicting extensions as shown in Example 1 since it may put two conflicting arguments in the same extension. One of these arguments is clearly undesirable. The second drawback is a consequence of the first one. Indeed, since an undesirable argument may be accepted, then all the arguments that are defended by this argument are accepted as well at the detriment of good ones. Let us illustrate this issue on an example.

Example 5 *(Cont).* Let us consider the following arguments of $\mathcal{F}_5$:

Expert: This violin is expensive since it was made by Stradivari (a).
Child_1: This violin was not made by Stradivari (b).
Child_2: This violin is very solid since it was made by Stradivari (c).
Child_1: The violin is not interesting since it is not solid (d).

The corresponding argumentation framework is depicted in the left side of the figure above. Assume that $a > b$. Using the PAF developed in [17] or the VAF introduced in [18], one gets the framework depicted in the right side of the same figure. Its grounded extension is the set $\{a, b, d\}$. This result is incorrect for two reasons: The first one is that child_1 and the expert cannot be both right. It is natural to have the argument of the expert as accepted while the argument b of the child as rejected. The second reason is that the argument b (which should be rejected) defends d against c, leading thus to an undesirable result. Indeed, d is defended by a "bad" argument! It is easy to check that our approach returns $\{a, c\}$ as the grounded extension and it rejects the two arguments of Child_1.

Our approach overcomes the limits of existing argumentation frameworks which deal with preferences or values. Furthermore, it is more general since it models even the second role of preferences (i.e., the refinement).

The first limit of existing models, namely the violation of conflict-freeness, has been pointed out in [39]. The authors proposed a new approach for handling critical attacks. It consists of defining a preference relation on the whole powerset of the set of arguments. The best elements of this relation are the extensions. The preference relation encodes the fact that an attack is privileged when it is not critical while the preference takes precedence when it is. Thus, that approach introduces preferences at the semantics level while ours does that at the level of the attack relation. Indeed, we repair this latter. It is worth mentioning that the approach in [39] neglects the refinement. To the best of our knowledge, there is only one work on refinement [23]. The authors proposed a particular refinement relation for symmetric argumentation frameworks [41] that use stable semantics. In this sense, our work is more general since it accepts any refinement relation. Moreover, there is no restriction to particular semantics or to particular attacks relations.

Modgil and Prakken proposed a preference-based argumentation framework in [42]. The framework is an instantiation of Dung's framework which considers preferences between arguments and three attack relations, namely rebutting, assumption attack and undercutting. This framework differs from our rich PAF in several aspects. First, it is an instantiation while our model is abstract. Second, it encodes only one role of preferences: handling critical attacks. Thus, it may return non-refined results. Our framework is thus richer since it captures both roles of preferences. Regarding the first role, the framework of Modgil and Prakken presents two limitations: First, a lot of complex conditions are assumed in order to ensure some very basic properties whereas our results hold for every argumentation framework. Second, the authors solved the problem of critical attacks only for symmetric relations like rebutting. However, undercutting may be in conflict with the preference relation and still always wins. Our approach is much more general and is suitable for both symmetric and non-symmetric attack relations.

8. Conclusion

This paper has presented a comprehensive study on the role that preferences can play in an argumentation framework. Two roles are distinguished. The first one consists of repairing the critical attacks. We have proposed a new approach for modeling this role and which overcomes the limitations of existing approaches. The basic idea is to invert the arrow of each critical attack instead of removing it. We have shown that such an approach is well-founded. The second role of preferences consists of refining the results of an argumentation framework. Indeed, we have shown that a refinement amounts to compare (using a preference relation, called a refinement relation) the extensions under a given semantics of an argumentation framework. It is clearly argued in the paper that the two roles are completely independent and should be modeled in different ways and at different steps of the evaluation process.

We have proposed the first abstract framework, called rich PAF, which models the two roles. The idea is to repair first the critical attacks, then to apply Dung's acceptability semantics on the repaired framework, and finally to apply a refinement relation on the extensions. In the general case, these steps are mandatory otherwise counter-intuitive results can be found.

We have also shown that the approach is well-founded since it allows to recover very well-known works on handling inconsistency in knowledge bases, namely the ones that restore consistency of knowledge bases. Indeed, we have shown full correspondences between instances of the new rich PAF and respectively the preferred sub-theories defined by Brewka in [29] and the democratic sub-theories proposed by Cayrol, Royer and Saurel in [30]. It is worth recalling that the same instance but without considering preferences was already used by Cayrol in [43] for making a bridge between stable extensions of an argumentation framework and the maximal consistent subsets of a propositional knowledge base.

Acknowledgements

The authors are very grateful to the anonymous referees for their valuable suggestions for improving the presentation of this work.

Appendix A

Proof of Property 1. Follows directly from Definition 9. $\square$

Proof of Property 2. Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$ be a PAF and $\mathcal{E}$ one of its extensions under one of the semantics recalled in Definition 5. Thus, $\mathcal{E}$ is conflict-free wrt. $\mathcal{R}_r$. Assume now that $\exists a, b \in \mathcal{E}$ such that $a\mathcal{R}b$. There are two cases:

- $b > a$: In this case $(b, a) \in \mathcal{R}_r$. This contradicts the fact that $\mathcal{E}$ is an extension, thus conflict-free.
- not $b > a$: In this case $(a, b) \in \mathcal{R}_r$. This contradicts the fact that $\mathcal{E}$ is an extension, thus conflict-free. $\square$

Proof of Property 3. Let $\mathcal{A}_a = \{x \in \mathcal{A} \mid a$ and x are related in $(\mathcal{A}, \mathcal{R})\}$. Let $\mathcal{R}_a = \{(x, y) \in \mathcal{R} \mid (x, y) \in \mathcal{A}_a \times \mathcal{A}_a\}$.

- Preferred semantics. We will first prove that for every preferred extension $\mathcal{E}$ of $(\mathcal{A}, \mathcal{R})$ there exists a preferred extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$ s.t. $\mathcal{E} \cap \mathcal{A}_a = \mathcal{E}' \cap \mathcal{A}_a$. Let $\mathcal{E}$ be a preferred extension of $(\mathcal{A}, \mathcal{R})$ and let $\mathcal{E}_a = \mathcal{E} \cap \mathcal{A}_a$. The set $\mathcal{E}_a$ is conflict-free and admissible in $(\mathcal{A}_a, \mathcal{R}_a)$. It is also admissible in $(\mathcal{A}, \mathcal{R}_r)$. Thus, there exists a preferred extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$ s.t. $\mathcal{E}_a \subseteq \mathcal{E}'$. Let $\mathcal{E}'_a = \mathcal{E}' \cap \mathcal{A}_a$. It holds that $\mathcal{E}_a \subseteq \mathcal{E}'_a$. Let us prove that $\mathcal{E}_a \subseteq \mathcal{E}'_a$. By means of contradiction, suppose that $\mathcal{E}_a \subsetneq \mathcal{E}'_a$. Thus, since $\mathcal{E}'_a$ is admissible in $(\mathcal{A}, \mathcal{R})$ and $(\mathcal{E} \setminus \mathcal{A}_a)$ is admissible in $(\mathcal{A}, \mathcal{R})$ and those two sets do not attack one another, then $\mathcal{E}'_a \cup (\mathcal{E} \setminus \mathcal{A}_a)$ is admissible in $(\mathcal{A}, \mathcal{R})$. Contradiction with the fact that $\mathcal{E}$ is a preferred extension in $(\mathcal{A}, \mathcal{R})$. By using exactly the same reasoning, we prove that for every preferred extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$ there exists a preferred extension $\mathcal{E}$ of $(\mathcal{A}, \mathcal{R})$ s.t. $\mathcal{E} \cap \mathcal{A}_a = \mathcal{E}' \cap \mathcal{A}_a$.
  Suppose now that $a \in \mathcal{A}_a$ is rejected in $(\mathcal{A}, \mathcal{R})$. This means that there is no preferred extension $\mathcal{E}$ of $(\mathcal{A}, \mathcal{R})$ such that $a \in \mathcal{E}$. From the previous fact, there is no preferred extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$ s.t. $a \in \mathcal{E}'$, so a must be rejected in $(\mathcal{A}, \mathcal{R}_r)$. The same reasoning yields the conclusion that the contrary is also true. By similar reasoning, we obtain that a is credulous in $(\mathcal{A}, \mathcal{R})$ iff a is credulous in $(\mathcal{A}, \mathcal{R}_r)$. From these two facts, we conclude that a is skeptically accepted in $(\mathcal{A}, \mathcal{R})$ iff it is skeptically accepted in $(\mathcal{A}, \mathcal{R}_r)$.
- Stable semantics. Throughout the proof, we suppose that there exists at least one stable extension in $(\mathcal{A}, \mathcal{R})$ and that there exists at least one stable extension in $(\mathcal{A}, \mathcal{R}_r)$. We will first prove that for every stable extension $\mathcal{E}$ of $(\mathcal{A}, \mathcal{R})$, there exists a stable extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$ s.t. $\mathcal{E} \cap \mathcal{A}_a = \mathcal{E}' \cap \mathcal{A}_a$.
  Let $\mathcal{E}$ be a stable extension of $(\mathcal{A}, \mathcal{R})$. Let $\mathcal{E}_a = \mathcal{E} \cap \mathcal{A}_a$. The set $\mathcal{E}_a$ is a stable extension of $(\mathcal{A}_a, \mathcal{R}_a)$. Let $\mathcal{E}_1$ be an arbitrary stable extension of $(\mathcal{A} \setminus \mathcal{A}_a, \mathcal{R}_r|_{\mathcal{A} \setminus \mathcal{A}_a})$ – it is easy to see that there must exist such an extension. Then, $\mathcal{E}_a \cup \mathcal{E}_1$ is a stable extension of $(\mathcal{A}, \mathcal{R}_r)$. To prove that for every stable extension $\mathcal{E}'$ of $(\mathcal{A}, \mathcal{R}_r)$, there exists a stable extension of $(\mathcal{A}, \mathcal{R})$ s.t. $\mathcal{E} \cap \mathcal{A}_a = \mathcal{E}' \cap \mathcal{A}_a$ is similar.
  By using this property, we can easily see that the status of an arbitrary argument $x \in \mathcal{A}_a$ is the same in $(\mathcal{A}, \mathcal{R})$ and $(\mathcal{A}, \mathcal{R}_r)$.
- Grounded semantics. Let, for an arbitrary set $S \subseteq \mathcal{A}$, $\mathcal{F}(S)_{(\mathcal{A},\mathcal{R})} = \{x \in \mathcal{A} \mid (\forall y \in \mathcal{A})$ if $y\mathcal{R}x$ then $(\exists z \in S)\ z\mathcal{R}y\}$. Let $\mathcal{F}^i(S)_{(\mathcal{A},\mathcal{R})} = \overbrace{\mathcal{F}(\mathcal{F}(\mathcal{F}(\ldots}^{i\text{ times}}(S)\ldots)_{(\mathcal{A},\mathcal{R})}$. We will prove by induction on i that $(\forall i \in \mathbb{N})\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R})} \cap \mathcal{A}_a = \mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R}_r)} \cap \mathcal{A}_a$.
  - Base. Follows from the fact that non-attacked arguments in $\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R})} \cap \mathcal{A}_a$ are exactly the non-attacked arguments in $\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R}_r)} \cap \mathcal{A}_a$.
  - Step. Let $\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R})} \cap \mathcal{A}_a = \mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R}_r)} \cap \mathcal{A}_a$. It is easy to see that an arbitrary argument in $\mathcal{A}_a$ is defended by $\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R})} \cap \mathcal{A}_a$ if and only if it is defended by $\mathcal{F}^i(\emptyset)_{(\mathcal{A},\mathcal{R}_r)} \cap \mathcal{A}_a$. Thus, $\mathcal{F}^{i+1}(\emptyset)_{(\mathcal{A},\mathcal{R})} \cap \mathcal{A}_a = \mathcal{F}^{i+1}(\emptyset)_{(\mathcal{A},\mathcal{R}_r)} \cap \mathcal{A}_a$.

  We conclude by induction that $(\forall i \in \mathbb{N})\ \mathcal{F}^i(\emptyset)(\mathcal{A}, \mathcal{R}) \cap \mathcal{A}_a = \mathcal{F}^i(\emptyset)(\mathcal{A}, \mathcal{R}_r) \cap \mathcal{A}_a$. Let us use the notation $\mathrm{GE}(\mathcal{A}, \mathcal{R})$ for the grounded extension of argumentation framework $(\mathcal{A}, \mathcal{R})$. It has been shown in [16] that $\mathrm{GE}(\mathcal{A}, \mathcal{R}) = \bigcup_{i=0}^{\infty} \mathcal{F}^i(\emptyset)(\mathcal{A}, \mathcal{R})$. Since the number of arguments is supposed to be finite, $(\exists j \in \mathbb{N})$ s.t. $\mathcal{F}^j(\emptyset)(\mathcal{A}, \mathcal{R}) = \mathcal{F}^{j+1}(\emptyset)(\mathcal{A}, \mathcal{R})$ and $\mathcal{F}^j(\emptyset)(\mathcal{A}, \mathcal{R}_r) = \mathcal{F}^{j+1}(\emptyset)(\mathcal{A}, \mathcal{R}_r)$. From those facts, we have that $\mathrm{GE}(\mathcal{A}, \mathcal{R}) \cap \mathcal{A}_a = \mathrm{GE}(\mathcal{A}, \mathcal{R}_r) \cap \mathcal{A}_a$. This means that status of arguments in $\mathcal{A}_a$ do not change. $\square$

Proof of Property 4. Let $\mathtt{Max}(\mathcal{A}, \geqslant)$ be conflict-free wrt. $\mathcal{R}$ and let a be an arbitrary element of $\mathtt{Max}(\mathcal{A}, \geqslant)$. We prove that a is not attacked wrt. $\mathcal{R}_r$. On the one hand, since $\mathtt{Max}(\mathcal{A}, \geqslant)$ is conflict-free, a could only be attacked (wrt. $\mathcal{R}_r$) by some argument b not belonging to $\mathtt{Max}(\mathcal{A}, \geqslant)$, i.e. from an argument b such that $a > b$. On the other hand, from the definition of $\mathcal{R}_r$, we see that there can exist no arguments $a, b \in \mathcal{A}$ s.t. $b\mathcal{R}_r a \wedge a > b$. So, a cannot be attacked by an argument b such that $a > b$. Those two facts imply that $(\nexists b \in \mathcal{A}) b\mathcal{R}_r a$.

- Since a is not attacked, it is in the grounded extension. Let $\mathcal{E}$ be a preferred extension. Suppose that $(\nexists b \in \mathcal{E}) a\mathcal{R}_r b$. In that case, it can easily be checked that since $\mathcal{E}$ and $\{a\}$ are both admissible sets that do not attack one another, then the set $\mathcal{E} \cup \{a\}$ is also admissible in $\mathcal{T}$. Contradiction with the definition of preferred extension. Suppose that $(\exists b \in \mathcal{E}) a\mathcal{R}_r b$. Since $\mathcal{E}$ is a preferred extension, $(\exists c \in \mathcal{E}) c\mathcal{R}_r a$. Contradiction. Thus, it must be that $a \in \mathcal{E}$, for every stable or preferred extension in $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geqslant)$.
- Let $\mathcal{E}$ be a stable extension of $\mathcal{T}$ and suppose that $a \notin \mathcal{E}$. Since $\mathcal{E}$ is stable, then it attacks wrt. $\mathcal{R}_r$ any argument which does not belong to $\mathcal{E}$. Contradiction with the fact that a is not attacked in $\mathcal{T}$. $\square$

Proof of Property 5. Let us consider the following algorithm:

```
input:
A:      set of arguments
R:      attack relation
>=:     preference relation

output:
in:     the only stable/preferred/grounded ext.
out:    rejected arguments wrt. those semantics

/* Put all arguments in und. */
in = {};
out = {};
und = A;

/*  While und is not empty,
sort arguments from und to in and out. */
while (not (und == {}) {

    /* Select the best argument in und,
    and move it to in. */
    let a be the only argument in the set
        {x in und | for all x' in und, x > x};
    in = in union {a};
    und = und - {a};

    /* Since a is accepted, all arguments being
    in conflict with it must be rejected. */
    del = {x in und | x R a  or  a R x};
    out = out union del;
    und = und - del;
}
```

Let us prove that in is a stable extension of $\mathcal{T}$. It is clear that in is conflict-free. Let $x' \notin in$. From the previous algorithm, it is easy to see that there exists $x \in in$ s.t. $x > x'$ and ($x\mathcal{R}x'$ or $x'\mathcal{R}x$). In other words, $x\mathcal{R}_r x'$. Thus, in is a stable extension of $\mathcal{T}$.

Every stable extension is a preferred and a complete extension [16]. Thus, in is a preferred and complete extension of $\mathcal{T}$.

Let us prove that in is the only complete extension. Suppose that $\mathcal{E} \subseteq \mathcal{A}$, with $\mathcal{E} \neq in$ is another complete extension. Since none of the arguments of in is attacked (wrt. $\mathcal{R}_r$), it is clear that every complete extension must contain those arguments, i.e., $in \subsetneq \mathcal{E}$. But, since in is a stable extension, it is maximal conflict-free set, contradiction. So, we have shown that in is the only complete extension.

Grounded extension is exactly the intersection of all complete extensions [16]. Hence, in is the grounded extension of $\mathcal{T}$.

Let us now prove that in is the only stable and the only preferred extension. Suppose not, thus there exists another stable or preferred extension $\mathcal{E}$, such that $\mathcal{E} \neq in$. Since we suppose that $\mathcal{E}$ is stable or preferred, then $\mathcal{E}$ is for sure complete [16]. But we have already shown that in was a unique complete extension, contradiction. Thus, in is the unique stable and preferred extension of $\mathcal{T}$.

The while loop is executed at most n times, where n is the number of arguments, and its execution contains at most n comparisons. Thus, algorithm's time complexity is $\mathcal{O}(n^2)$. $\square$

Proof of Property 6. It is easy to see that when $\mathcal{R}$ is symmetric then $\mathcal{R}_r = \mathcal{R}'$. Thus, extensions of the two frameworks must coincide as well. $\square$

Proof of Property 7. It is easy to see that a set is a stable extension of $(\mathcal{A}, \mathcal{R})$ iff it is a preferred extension of $(\mathcal{A}, \mathcal{R})$ iff it is a maximal conflict-free set, since every maximal conflict-free set attacks all arguments in its exterior in this framework. Let us now suppose that $\mathcal{E}$ is a preferred (stable) extension of $\mathcal{T}$. This means that $\mathcal{E}$ is maximal conflict-free set in $\mathcal{T}$. It is immediate that $\mathcal{E}$ is maximal conflict-free in $(\mathcal{A}, \mathcal{R})$. $\square$

Proof of Property 8. Let $\mathcal{E} \subseteq \mathcal{A}$ be a stable extension of $(\mathcal{A}, \mathcal{R})$ which is not a stable extension of $\mathcal{T}$. This can be true iff $\mathcal{E}$ is a maximal conflict-free set but that it is not stable in $\mathcal{T}$. Formally, $(\exists x' \notin \mathcal{E})$ $(\nexists x \in \mathcal{E}) x\mathcal{R}_r x'$. This is equivalent to $(\exists x' \notin \mathcal{E})$ $(\forall x \in \mathcal{E})\neg(x\mathcal{R}_r x')$. It is obvious that $\neg(x\mathcal{R}_r x')$ is equivalent to $x\mathcal{R}_r x' \Rightarrow x' > x$. $\square$

Proof of Property 9. Property 6 shows that in the case when $\mathcal{R}$ is symmetric, extensions of $\mathcal{T}$ coincide with extensions of framework $(\mathcal{A}, \mathcal{R}')$, where $\mathcal{R}'$ is defined as in Property 6. Theorems 1 and 2 from [44] imply that $\mathcal{T}$ is coherent and that it has at least one stable extension. $\square$

Proof of Property 10. We will first study the democratic relation $\succcurlyeq_d$.

- From the definition of democratic relation we see that it is reflexive.
- Let us prove that democratic order is a transitive relation. Let $\mathcal{X} \succcurlyeq_d \mathcal{Y}$ and $\mathcal{Y} \succcurlyeq_d \mathcal{Z}$. We will prove that $\mathcal{X} \succcurlyeq_d \mathcal{Z}$. Let $z \in \mathcal{Z} \setminus \mathcal{X}$. Let us study two possible cases:
  - Let $z \notin \mathcal{Y}$. Since $\mathcal{Y} \succcurlyeq_d \mathcal{Z}$ then $(\exists y \in \mathcal{Y} \setminus \mathcal{Z})$ s.t. $y > z$. If $y \in \mathcal{X}$ then the proof is over. If it is not the case, then $y \in \mathcal{Y} \setminus \mathcal{X}$. So, $(\exists x \in \mathcal{X} \setminus \mathcal{Y})$ s.t. $x > y$. If $x \notin \mathcal{Z}$ then the proof is over. If $x \in \mathcal{X}$, $x \in \mathcal{Z}$ and $x \notin \mathcal{Y}$ then, since $x \in \mathcal{Z}$ we have that $(\exists y_1 \in \mathcal{Y} \setminus \mathcal{Z})$ s.t. $y_1 > x$. If $y_1 \in \mathcal{X}$, then we have $y_1 > x > y > z$ and the proof is over. If that is not the case, it must be that $y_1 \notin \mathcal{X}$. Thus, $(\exists x_1 \in \mathcal{X} \setminus \mathcal{Y})$ s.t. $x_1 > y_1$. If $x_1 \notin \mathcal{Z}$, from the transitivity of preference relation, we have $x_2 > z$. In the case when $x_1 \in \mathcal{Z}$, since it is not in $\mathcal{Y}$, we have that $(\exists y_2 \in \mathcal{Y} \setminus \mathcal{Z})$ s.t. $y_2 > x_1$. From $y_2 > y_1$, we have that $y_2 \neq y_1$. Either we will end the proof or we continue by constructing an infinite sequence of different arguments $y_1, \dots, y_n, \dots$ while we supposed that $\mathcal{Y}$ is finite. Contradiction.
  - Let $z \notin \mathcal{X}$, $z \in \mathcal{Y}$, $z \in \mathcal{Z}$. If $(\exists x \in \mathcal{X} \setminus \mathcal{Z})$ s.t. $x > z$, the proof is over. Else, $\exists x$ s.t. $x \in \mathcal{X} \cap \mathcal{Z}$, $x \notin \mathcal{Y}$, $x > z$. Then, it must be that $(\exists y \in \mathcal{Y} \setminus \mathcal{Z})$ $y > x$. From the previous facts, we have also that $y \notin \mathcal{X}$, since in that case, from the transitivity of preference relation, we have $y > z$, contradiction. Thus, $y \notin \mathcal{X}$, $y \in \mathcal{Y}$, $y \notin \mathcal{Z}$. From all those facts and from $\mathcal{X} \succcurlyeq_d \mathcal{Y}$, $\exists x_1 \in (\mathcal{X} \cap \mathcal{Z}) \setminus \mathcal{Y}$ s.t. $x_1 > y$. Similarly, $\exists y_1 \in \mathcal{Y} \setminus (\mathcal{X} \cup \mathcal{Z})$ s.t. $y_1 > x_1$. It is obvious how an infinite sequence of different arguments can be constructed, despite the fact that we supposed that the set $\mathcal{A}$, and, consequently, the sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are finite. Contradiction.
- The third property required by the definition of refinement relation is trivially satisfied by $\succcurlyeq_d$.

Now, we prove that $\succcurlyeq_e$ is also a refinement relation.

- It is trivial that the relation $\succcurlyeq_e$ is reflexive.
- Let us prove that this relation is transitive. Let $\mathcal{X} \succcurlyeq_e \mathcal{Y}$ and $\mathcal{Y} \succcurlyeq_e \mathcal{Z}$ and let us prove that $\mathcal{X} \succcurlyeq_e \mathcal{Z}$ must hold. Let $x \in \mathcal{X} \setminus \mathcal{Z}$. We study two cases:
  - $x \in \mathcal{Y}$. In this case, from $\mathcal{Y} \succcurlyeq_e \mathcal{Z}$, it holds that $(\exists z' \in \mathcal{Z} \setminus \mathcal{Y})$ s.t. $x > z'$. If $(\exists z' \in \mathcal{Z} \setminus \mathcal{Y})$ s.t. $x > z'$ and $z' \notin \mathcal{X}$, the proof is over. We now study the second possible case, the one when $(\nexists z' \in \mathcal{Z} \setminus (\mathcal{X} \cup \mathcal{Y}))$ s.t. $x > z'$. Thus, $(\exists z_1 \in \mathcal{Z} \setminus \mathcal{Y})$ s.t. $x > z'$ and $z_1 \in \mathcal{X}$. Since $z_1 \in \mathcal{X} \setminus \mathcal{Y}$ and $X \succcurlyeq_e \mathcal{Y}$, then $(\exists y_1 \in \mathcal{Y} \setminus \mathcal{X})$ s.t. $z_1 > y_1$. Note that transitivity of preference relation implies that $x > y_1$. If $y_1 \in \mathcal{Z}$ the proof is over. Else, let $y_1 \notin \mathcal{Z}$. Thus, we have $y_1 \in \mathcal{Y} \setminus (\mathcal{X} \cup \mathcal{Z})$. From $\mathcal{Y} \succcurlyeq_e \mathcal{Z}$, we obtain $\exists z_2 \in \mathcal{Z} \setminus \mathcal{Y}$ s.t. $y_1 > z_2$. If $z_2 \notin \mathcal{X}$, then the fact that $x > z_2$ ends the proof, since it means that $\mathcal{X} \succcurlyeq_e \mathcal{Z}$. Else, $z_2 \in \mathcal{X}$. By following this idea, we will either find an element $z \in \mathcal{Z} \setminus \mathcal{X}$ s.t. $x > z$, which will end the proof or we will construct the chain $z_1 > z_2 > z_3 > \cdots$ which contains infinitely many arguments. Contradiction with the fact that $\mathcal{A}$ is finite.
  - $x \notin \mathcal{Y}$. In this case, from $\mathcal{X} \succcurlyeq_e \mathcal{Y}$, we have that $(\exists y \in \mathcal{Y} \setminus \mathcal{X})$ s.t. $x > y$. If $y \in \mathcal{Z}$, the proof is over, since we obtain $\mathcal{X} \succcurlyeq_e \mathcal{Z}$. Else, let $y \notin \mathcal{Z}$. Now, the rest of the proof is similar to the proof of previous item.
- It is easy to see that the third item of Definition 14 is satisfied in the case of $\succcurlyeq_e$. $\square$

Proof of Property 11. It is easy to see why the first and the second items of this property hold. The third one follows from Property 5. $\square$

Proof of Property 14. Let $\mathcal{S}$ be a democratic sub-theory. From Definition 17, $\mathcal{S}$ is consistent. Assume now that $\mathcal{S}$ is not a maximal (for set inclusion) consistent set. Thus, $\exists x \in \Sigma \setminus \mathcal{S}$ s.t. $\mathcal{S} \cup \{x\}$ is consistent. It is clear that $\mathcal{S} \cup \{x\} \succ_d \mathcal{S}$. This contradicts the fact that $\mathcal{S}$ is a democratic sub-theory. $\square$

Proof of Property 15. Let $\mathcal{E}, \mathcal{E}'$ be two stable extensions of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant)$, and let $\mathcal{E} \succcurlyeq_d \mathcal{E}'$ with $\mathcal{E} \neq \mathcal{E}'$. It is clear that $\neg(\mathcal{E} \subseteq \mathcal{E}')$ and $\neg(\mathcal{E}' \subseteq \mathcal{E})$. Let $a' \in \mathcal{E}' \setminus \mathcal{E}$ be such that $\forall a'' \in \mathcal{E}' \setminus \mathcal{E}$ it holds that $a' \geqslant a''$ (this is possible since $\geqslant$ is a total preorder). From $\mathcal{E} \succcurlyeq_d \mathcal{E}'$, we have that $\exists a \in \mathcal{E} \setminus \mathcal{E}'$ s.t. $a > a'$. This means that $\forall b' \in \mathcal{E}' \setminus \mathcal{E}$, $a > b'$. Since $\mathcal{E}'$ is a stable extension, then $\exists a' \in \mathcal{E}'$ s.t. $a' \mathcal{R}_r a$, i.e. $(a' \mathcal{R} a$ and $\neg(a > a'))$ or $(a \mathcal{R} a'$ and $a' > a)$. Sets $\mathcal{E}$ and $\mathcal{E}'$ are both conflict-free, so $a' \in \mathcal{E}' \setminus \mathcal{E}$. Contradiction, since $\forall a' \in \mathcal{E}' \setminus \mathcal{E}$ we have $a > a'$. $\square$

Proof of Property 18. Let $\mathcal{S} \subseteq \Sigma$.

- Assume that $\mathcal{S}$ is consistent and $\mathtt{Arg}(\mathcal{S})$ is not conflict-free. This means that there exist $a, a' \in \mathtt{Arg}(\mathcal{S})$ s.t. a undercuts a'. From Definition 2 of undercut, it follows that $\mathtt{Supp}(a) \cup \mathtt{Supp}(a')$ is inconsistent. Besides, from the definition of an argument, $\mathtt{Supp}(a) \subseteq \mathcal{S}$ and $\mathtt{Supp}(a') \subseteq \mathcal{S}$. Thus, $\mathtt{Supp}(a) \cup \mathtt{Supp}(a') \subseteq \mathcal{S}$. Then, $\mathcal{S}$ is inconsistent. Contradiction.
- Assume now that $\mathcal{S}$ is inconsistent. This means that there exists a finite set $\mathcal{S}' = \{h_1, \ldots, h_k\}$ s.t.
  - $\mathcal{S}' \subseteq \mathcal{S}$
  - $\mathcal{S}' \vdash \bot$
  - $\mathcal{S}'$ is minimal (wrt. set inclusion) s.t. previous two items hold.

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

Proof of Theorem 1. Let $\mathcal{S}$ be a preferred sub-theory of a knowledge base Σ. Thus, $\mathcal{S}$ is consistent. From Property 18, it follows that $\mathtt{Arg}(\mathcal{S})$ is conflict-free. Assume that $a \notin \mathtt{Arg}(\mathcal{S})$. Since $a \notin \mathtt{Arg}(\mathcal{S})$ and $\mathcal{S}$ is a maximal consistent subbase of Σ (according to Property 13), then $\exists h \in \mathtt{Supp}(a)$ s.t. $\mathcal{S} \cup \{h\} \vdash \bot$. Assume that $h \in \Sigma_j$. Thus, $\mathtt{Level}(\mathtt{Supp}(a)) \geqslant j$.

Since $\mathcal{S}$ is a preferred sub-theory of Σ, then $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_j$ is a maximal (for set inclusion) consistent subbase of $\Sigma_1 \cup \cdots \cup \Sigma_j$. Thus, $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_j \cup \{h\} \vdash \bot$. This means that there exists an argument $a' = (\mathcal{S}', \neg h)$ s.t. $a' \in \mathtt{Arg}(\mathcal{S})$ and $\mathcal{S}' \subseteq \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_j$. Thus, $\mathtt{Level}(\mathcal{S}') \leqslant j$. Consequently, $a' \geqslant_{wlp} a$. From this fact, together with $a' \mathcal{R} a$, we obtain $a' R_r a$. $\square$

The second part of the theorem follows directly from Property 17.

Proof of Theorem 2. Let Σ be a stratified knowledge base. Throughout the proof, we will use the notation $\mathcal{S}_i = \mathcal{S} \cap \Sigma_i$ and $\mathtt{PST}(\Sigma)$ denotes the set of preferred sub-theories of Σ. Let $\mathcal{E}$ be a stable extension of $(\mathtt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$. Let $\mathcal{S} = \mathtt{Base}(\mathcal{E})$.

- We will first show that $\mathcal{S} \in \mathtt{PST}(\Sigma)$. Suppose that $\mathcal{S} \notin \mathtt{PST}(\Sigma)$. If $\mathcal{S}$ is not consistent, then Property 18 implies that $\mathcal{E}$ is not conflict-free. This contradicts the fact that $\mathcal{E}$ is a stable extension. Thus $\mathcal{S}$ is consistent but it is not a preferred sub-theory. Thus, there exists $i \in \{1, \ldots, n\}$ such that $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_i$ is not a maximal consistent set in $\Sigma_1, \ldots, \Sigma_i$. Let i be minimal s.t. $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_i$ is not a maximal consistent set in $\Sigma_1, \ldots, \Sigma_i$. This means that there exists $x \notin \mathcal{S}$ s.t. $x \in \Sigma_i$ and $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_i \cup \{x\}$ is consistent. Thus, $a' = (\{x\}, x)$ is an argument. Thus, $(\exists a \in \mathcal{E})$ s.t. $a \mathcal{R}_r a'$. Since $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_i \cup \{x\}$ 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 $\mathcal{E} = \mathtt{Arg}(\mathcal{S})$. Suppose the contrary. From Property 3, $\mathcal{E} \subseteq \mathtt{Arg}(\mathtt{Base}(\mathcal{E}))$, thus $\mathcal{E} \subsetneq \mathtt{Arg}(\mathtt{Base}(\mathcal{E}))$.
  - Let us suppose that $\mathcal{S}$ is consistent. Since $\mathcal{S}$ is consistent, then Property 18 implies that $\mathtt{Arg}(\mathcal{S})$ is conflict-free. Since we supposed that $\mathcal{E} \subsetneq \mathtt{Arg}(\mathcal{S})$, 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 $\mathcal{S}' = \{h'_1, \ldots, h'_k\}$ s.t.
    * $\mathcal{S}' \subseteq \mathcal{S}$
    * $\mathcal{S}' \vdash \bot$
    * $\mathcal{S}'$ is a minimal s.t. the previous two conditions are satisfied.

    Let us consider the set $\mathcal{E}'$ containing the following k arguments: $\mathcal{E}' = \{a'_1, \ldots, a'_k\}$, 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. $h'_i \in \mathtt{Supp}(a)$ and since $\mathcal{E}$ is conflict-free then $(\nexists b \in \mathcal{E})$ s.t. $\mathtt{Conc}(b) \in \{\neg h'_1, \ldots, \neg h'_k\}$. 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 $a'_1, \ldots, a'_k$. 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:

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

For undercuts to exist, it is necessary that:

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

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

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

Let $l_i = \texttt{Level}(h'_i)$, 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:

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

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 $\{l_1,\dots,l_k\}$. □

Proof of Theorem 3. Let $\mathcal{T} = (\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{WLP})$ be an argumentation framework built over a stratified knowledge base Σ. Let $\texttt{PST}(\Sigma)$ denote the set of preferred sub-theories of Σ.

- For all $\mathcal{S} \in \texttt{PST}(\Sigma)$, Theorem 1 shows that $\texttt{Arg}(\mathcal{S}) \in \texttt{Ext}(\mathcal{T})$.
- Property 3 implies that $\texttt{Arg}$ is injective.
- Let $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$ and let $\mathcal{S} = \texttt{Base}(\mathcal{E})$. From Theorem 2, we have $\mathcal{E} = \texttt{Arg}(\mathcal{S})$. Theorem 2 yields the conclusion that $\mathcal{S} \in \texttt{PST}(\Sigma)$. Thus, $\texttt{Arg} : \texttt{PST}(\Sigma) \to \texttt{Ext}(\mathcal{T})$ is surjective. □

Proof of Theorem 4. Let us denote the set of democratic sub-theories by $\texttt{DMS}(\Sigma)$. Let $\mathcal{E} = \texttt{Arg}(\mathcal{S})$ and let $x \rhd x'$ iff $x \unrhd x'$ and not $x' \unrhd x$. From Property 18, we see that $\mathcal{E}$ is conflict-free. We will prove that it attacks (wrt. $\mathcal{R}_r$) any argument in its exterior. Let $a' \in \mathcal{A} \setminus \mathcal{E}$ be an arbitrary argument. Since $a' \notin \mathcal{E}$ then $\exists h' \in \texttt{Supp}(a')$ s.t. $h' \notin \mathcal{S}$. From $\mathcal{S} \in \texttt{DMS}(\Sigma)$ we have that $\mathcal{S}$ is a maximal consistent set. It is clear that $\mathcal{S} \cup \{h'\} \vdash \perp$. Let us identify all its minimal conflicting subsets. Formally, let $C_1,\dots,C_k$ be all sets which satisfy the following three conditions:

1. $C_i \subseteq \mathcal{S}$.
2. $C_i \cup \{h'\} \vdash \perp$.
3. C_i is minimal (wrt. set inclusion) s.t. the two previous conditions are satisfied.

Those sets allow to construct the following k arguments: $a_1 = (C_1, \neg h'),\dots,a_k = (C_k, \neg h)$. It is obvious that all of them attack a' wrt. $\mathcal{R}$. If at least one of them attack a' wrt. $\mathcal{R}_r$, then the proof is over. Suppose the contrary. This would mean that $\forall i \in \{1,\dots,k\}$, $a' > a_i$. Thus, $(\forall i \in \{1,\dots,k\})$ $(\exists h_i \in C_i)$ s.t. $h' \rhd h_i$. In other words, for every argument a_i, there exists one formula $h_i \in \texttt{Supp}(a_i)$, such that $h' \rhd h_i$. Let $H = \{h_1,\dots,h_k\}$.

Now, we can define a set $\mathcal{S}'$ as follows: $\mathcal{S}' = \mathcal{S} \cup \{h'\} \setminus H$. We will show that $\mathcal{S}'$ is consistent. Suppose the contrary. Since $\mathcal{S}$ is consistent, then any inconsistent subset of $\mathcal{S}'$ must contain h'. Let $K_1,\dots,K_j$ be all sets which satisfy the following conditions:

1. $K_i \subseteq \mathcal{S}' \setminus \{h'\}$.
2. $K_i \cup \{h'\} \vdash \perp$.
3. K_i is a minimal set s.t. the previous two conditions hold.

Let $K = \{K_1,\dots,K_j\}$ and $C = \{C_1,\dots,C_k\}$. It is easy to see that $K \subseteq C$ (this follows immediately from the fact that $\mathcal{S}' \setminus \{h'\} \subseteq \mathcal{S}$). Furthermore, since $(\forall C_i \in C)$ $(\exists h \in H)$ s.t. $h \in C_i$ then $(\forall K_i \in K)$ $(\exists h \in H)$ s.t. $h \in K_i$. Since for all K_i, we have that $K_i \cap H = \emptyset$ then it must be that $j = 0$, i.e. $K = \emptyset$. In other words, there are no inconsistent subsets of S', which means that S' is consistent.

We can notice that $\mathcal{S}' \setminus \mathcal{S} = \{h'\}$ and $\mathcal{S} \setminus \mathcal{S}' = \{h_1,\dots,h_k\}$. Since $\mathcal{S}'$ is consistent, we see that $\mathcal{S}' \succ \mathcal{S}$. Contradiction with $\mathcal{S} \in \texttt{DMS}(\Sigma)$. □

Proof of Theorem 5. Let $\mathcal{T} = (\texttt{Arg}(\Sigma), \text{Undercut}, \geqslant_{GWLP})$ be an argumentation framework built over a knowledge base Σ. Let $\mathcal{E}$ be a stable extension of $\mathcal{T}$ and let $\mathcal{S} = \texttt{Base}(\mathcal{E})$.

- Let us suppose that $\mathcal{S}$ is consistent but that it is not a maximal consistent set. This means that $\exists h \in \Sigma \setminus \mathcal{S}$ s.t. $\mathcal{S} \cup \{h\}$ is consistent. From Property 18, $\mathcal{E}' = \texttt{Arg}(\mathcal{S} \cup \{h\})$ is consistent. From Property 3, $\mathcal{E} \subseteq \mathcal{E}'$. The same result implies that $\mathcal{E} \neq \mathcal{E}'$. Thus, $\mathcal{E} \subsetneq \mathcal{E}'$, which means that $\mathcal{E}$ is not a maximal conflict-free set. Contradiction with the fact that $\mathcal{E}$ is a stable extension.
- Suppose now that $\mathcal{S}$ is inconsistent. This means that there can be found a set $S' = \{h'_1, \ldots, h'_k\}$ 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: $\mathcal{E}' = \{a'_1, \ldots, a'_k\}$, 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. $h'_i \in \texttt{Supp}(a)$ and since $\mathcal{E}$ is conflict-free then $(\nexists b \in \mathcal{E})$ s.t. $\texttt{Conc}(b) \in \{\neg h'_1, \ldots, \neg h'_k\}$. 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 $a'_1, \ldots, a'_k$. 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:

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

  For undercuts to exist, it is necessary that:

  $$\big(\forall i \in \{1, \ldots, k\}\big)\quad \big(h'_i \in \texttt{Supp}(a_i)\big) \wedge \big(a_i > a'_i\big).$$

  For $i = 1$, we have: $\exists i_1 \in \{1, \ldots, k\}$ s.t. $h_1 \rhd h_{i_1}$. For $i = i_1$, we have that $\exists i_2 \in \{1, \ldots, k\}$ s.t. $h_{i_1} \rhd h_{i_2}$, thus, $h_1 \rhd h_{i_1} \rhd h_{i_2}$. After k consecutive applications of the same rule, we obtain: $h_1 \rhd h_{i_1} \rhd \cdots \rhd h_{i_k}$. It is clearly a contradiction since on one hand, all the formulae in the chain are different because of the strict preference between them, and, on the other hand, set $\{h_1, \ldots, h_k\}$ contains k formulae, thus at least two of them in a chain of $k+1$ formulae must coincide.

This ends the first part of the proof. Let us now prove that S is a maximal consistent set in Σ. From Property 3, we have that $\mathcal{E} \subseteq \texttt{Arg}(\mathcal{S})$. Suppose that $\mathcal{E} \subsetneq \texttt{Arg}(\mathcal{S})$. Property 5 implies that $\mathcal{S}$ is a maximal consistent set. Thus, from Property 18, we have that $\texttt{Arg}(\mathcal{S})$ is conflict-free. This simply means that $\mathcal{E}$ is not a maximal conflict-free set, contradiction. □

Proof of Theorem 6. Let $\mathcal{S}, \mathcal{S}' \subseteq \Sigma$ be maximal (for set inclusion) consistent subbases of Σ.

($\Rightarrow$) Let $\mathcal{S} \succcurlyeq_d \mathcal{S}'$. Let $a' \in \mathcal{E}' \setminus \mathcal{E}$. Then $\exists h' \in \texttt{Supp}(a')$ s.t. $h' \in \mathcal{S}' \setminus \mathcal{S}$. Since $\mathcal{S} \succcurlyeq_d \mathcal{S}'$ then $\exists h \in \mathcal{S} \setminus \mathcal{S}'$ s.t. $h \rhd h'$. Let $a = (\{h\}, h)$. It is clear that $a \in \mathcal{S} \setminus \mathcal{S}'$ and $a > a'$. Thus, $\mathcal{E} \succcurlyeq_d \mathcal{E}'$.

($\Leftarrow$) Let $\mathcal{E} \succcurlyeq_d \mathcal{E}'$. Let $h' \in \mathcal{S}' \setminus \mathcal{S}$. Then $a' = (\{h'\}, h') \in \mathcal{E}' \setminus \mathcal{E}$. Thus, $\exists a \in \mathcal{E} \setminus \mathcal{E}'$ s.t. $a > a'$. Since $a \in \mathcal{E} \setminus \mathcal{E}'$, then $\exists h \in \texttt{Supp}(a)$ s.t. $h \in \mathcal{S} \setminus \mathcal{S}'$. It is clear that $h \rhd h'$. □

Proof of Theorem 7. Let us denote the set of democratic sub-theories by $\texttt{DMS}(\Sigma)$.

- From Theorem 4, we have that $\mathcal{E}$ is an extension of a basic PAF $(\mathcal{A}, \mathcal{R}, \geqslant)$. We will prove that it is also an extension of a rich PAF $(\mathcal{A}, \mathcal{R}, \geqslant, \succcurlyeq_d)$. Let us suppose the contrary, i.e. suppose that there exists $\mathcal{E}'$ s.t. $\mathcal{E}'$ is a stable extension and $\mathcal{E}' \succ_d \mathcal{E}$. Let $\mathcal{S}' = \texttt{Base}(\mathcal{E}')$. From Property 5, we have that $\mathcal{S}'$ is maximal consistent set and from Property 6 that $\mathcal{S}' \succ_d \mathcal{S}$. Contradiction.
- Property 5 implies that $\mathcal{S}$ is a maximal conflict-free set. Suppose that $\mathcal{S} \notin \texttt{DMS}(\Sigma)$. This means that $\exists \mathcal{S}' \subseteq \Sigma$ s.t. $\mathcal{S}' \in \texttt{DMS}(\Sigma)$ and $\mathcal{S}' \succ_d \mathcal{S}$. From Theorem 4, $\mathcal{E}' = \texttt{Arg}(\mathcal{S}')$ is a stable extension of a basic PAF. Property 6 implies that $\mathcal{E}' \succ_d \mathcal{E}$, contradiction. □

Proof of Theorem 8. Let us denote $\texttt{Ext}(\mathcal{T})$ the set of all extensions of a rich PAF $\mathcal{T}$ and $\texttt{DMS}(\Sigma)$ the set of democratic sub-theories of Σ. We will prove that $\texttt{Arg} : \texttt{DMS}(\Sigma) \to \texttt{Ext}(\mathcal{T})$ is a bijection.

- Let $\mathcal{S} \in \texttt{DMS}(\Sigma)$, Theorem 7 shows that $\texttt{Arg}(\mathcal{S}) \in \texttt{Ext}(\mathcal{T})$.
- Property 3 implies that $\texttt{Arg}$ is injective.
- Let $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$ and let $\mathcal{S} = \texttt{Base}(\mathcal{E})$. From Theorem 5, we have $\mathcal{E} = \texttt{Arg}(\mathcal{S})$. Theorem 7 yields the conclusion that $\mathcal{S} \in \texttt{DMS}(\Sigma)$. Thus, $\texttt{Arg} : \texttt{DMS}(\Sigma) \to \texttt{Ext}(\mathcal{T})$ is surjective. □

References

[1] L. Amgoud, S. Vesic, Handling inconsistency with preference-based argumentation, in: A. Deshpande, A. Hunter (Eds.), Proceedings of the 4th International Conference on Scalable Uncertainty Management, SUM'10, Springer, 2010, pp. 56–69.

[2] L. Amgoud, S. Vesic, On the role of preferences in argumentation frameworks, in: Proceedings of the 22nd International Conference on Tools in AI, ICTAI'10, IEEE, 2010, pp. 219–222.

[3] L. Amgoud, S. Vesic, Two roles of preferences in argumentation frameworks, in: W. Liu (Ed.), Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'11, Springer, 2011, pp. 86–97.

[4] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artif. Intell. 128 (2001) 203–235.

[5] A. García, G. Simari, Defeasible logic programming: an argumentative approach, Theory Pract. Log. Program. 4 (1) (2004) 95–138.

[6] G. Simari, R. Loui, A mathematical treatment of defeasible reasoning and its implementation, Artif. Intell. 53 (1992) 125–157.
[7] L. Amgoud, S. Kaci, An argumentation framework for merging conflicting knowledge bases, Int. J. Approx. Reason. 45 (2007) 321–340.
[8] L. Amgoud, H. Prade, Using arguments for making and explaining decisions, Artif. Intell. 173 (3–4) (2009) 413–436.
[9] B. Bonet, H. Geffner, Arguing for decisions: A qualitative model of decision making, in: Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, UAI'96, 1996, pp. 98–105.
[10] J. Müller, A. Hunter, An argumentation-based approach for decision making, in: 24th International Conference on Tools with Artificial Intelligence, ICTAI'12, 2012, pp. 564–571.
[11] L. Amgoud, S. Belabbes, H. Prade, Towards a formal framework for the search of a consensus between autonomous agents, in: Argumentation in Multi-Agent Systems, 2006, pp. 264–278.
[12] K. Sycara, Persuasive argumentation in negotiation, Theory Decis. 28 (1990) 203–242.
[13] P. Mcburney, E. Van, S. Parsons, L. Amgoud, A dialogue game protocol for agent purchase negotiations, Auton. Agents Multi-Agent Syst. 7 (3) (2003) 235–273.
[14] X. Fan, F. Toni, Argumentation dialogues for two-agent conflict resolution, in: Computational Models of Argument, COMMA'12, 2012, pp. 249–260.
[15] S. Zabala, I. Lara, H. Geffner, Beliefs, reasons and moves in a model for argumentation dialogues, in: Proceedings of the Latino-American Conference on Computer Science, 1999.
[16] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–357.
[17] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, Ann. Math. Artif. Intell. 34 (2002) 197–216.
[18] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, J. Log. Comput. 13 (3) (2003) 429–448.
[19] S. Kaci, L. van der Torre, Preference-based argumentation: Arguments supporting multiple values, Int. J. Approx. Reason. 48 (3) (2008) 730–751.
[20] S. Modgil, Reasoning about preferences in argumentation frameworks, Artif. Intell. 173 (9–10) (2009) 901–934.
[21] L. Amgoud, C. Cayrol, Inferring from inconsistency in preference-based argumentation frameworks, Int. J. Autom. Reason. 29 (2) (2002) 125–169.
[22] H. Prakken, G. Sartor, Argument-based extended logic programming with defeasible priorities, J. Appl. Non-Class. Log. 7 (1997) 25–75.
[23] Y. Dimopoulos, P. Moraitis, L. Amgoud, Extending argumentation to make good decisions, in: ADT'09, in: Lect. Notes Comput. Sci., vol. 5783, 2009, pp. 225–236.
[24] G. Brewka, T. Eiter, Preferred answer sets for extended logic programs, Artif. Intell. 109 (1–2) (1999) 297–356.
[25] G. Brewka, I. Niemelä, M. Truszczynski, Preferences and nonmonotonic reasoning, AI Mag. 29 (4) (2008) 69–78.
[26] R. Reiter, A logic for default reasoning, Artif. Intell. 13 (1980) 81–132.
[27] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artif. Intell. 171 (5–6) (2007) 286–310.
[28] S. Barbera, W. Bossert, P.K. Pattanaik, Ranking sets of objects, Technical report, Université de Montréal, Département de sciences économiques, 2001.
[29] G. Brewka, Preferred subtheories: An extended logical framework for default reasoning, in: Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, 1989, pp. 1043–1048.
[30] C. Cayrol, V. Royer, C. Saurel, Management of preferences in assumption-based reasoning, in: Lect. Notes Comput. Sci., vol. 682, 1993, pp. 13–22.
[31] L. Amgoud, S. Vesic, Repairing preference-based argumentation systems, in: 21st International Joint Conference on Artificial Intelligence, IJCAI'09, 2009, pp. 665–670.
[32] M. Elvang-Gøransson, J. Fox, P. Krause, Acceptability of arguments as 'logical uncertainty', in: Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU'93, vol. 747, Springer-Verlag, 1993, pp. 85–90.
[33] N. Gorogiannis, A. Hunter, Instantiating abstract argumentation with classical logic arguments: Postulates and properties, Artif. Intell. 175 (9–10) (2011) 1479–1497.
[34] S. Benferhat, D. Dubois, H. Prade, Argumentative inference in uncertain and inconsistent knowledge bases, in: Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence, 1993, pp. 411–419.
[35] L. Amgoud, Contribution à l'intégration des préférences dans le raisonnement argumentatif, Ph.D. thesis, Université Paul Sabatier, Toulouse, July 1999.
[36] L. Amgoud, Postulates for logic-based argumentation systems, Int. J. Approx. Reason. (2013), http://dx.doi.org/10.1016/j.ijar.2013.10.004, in press.
[37] L. Amgoud, P. Besnard, Bridging the gap between abstract argumentation systems and logic, in: International Conference on Scalable Uncertainty Management (SUM), Springer, Washington, DC, USA, 2009, pp. 12–27.
[38] A. Tarski, On some fundamental concepts of metamathematics, in: E.H. Woodger (Ed.), Logic, Semantics, Metamathematics, Oxford Univ. Press, 1956.
[39] L. Amgoud, S. Vesic, A new approach for preference-based argumentation frameworks, Ann. Math. Artif. Intell. 63 (2) (2011) 149–183.
[40] N. Rescher, R. Manor, On inference from inconsistent premises, Theory Decis. 1 (1970) 179–219.
[41] S. Coste, C. Devred, P. Marquis, Symmetric argumentation frameworks, in: 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'05, 2005, pp. 317–328.
[42] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artif. Intell. 195 (2013) 361–397.
[43] C. Cayrol, On the relation between argumentation and non-monotonic coherence-based entailment, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, 1995, pp. 1443–1448.
[44] L. Amgoud, Y. Dimopoulos, P. Moraitis, Making decisions through preference-based argumentation, in: Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning, KR'08, AAAI Press, 2008, pp. 113–123.